

66250

From: Chan, Christina
Sent: Wednesday, May 08, 2002 11:04 AM
To: Davis, Minh-Tam; STIC-Biotech/ChemLib
Subject: RE: Rush search request for 09/802520

Please rush. Thanks Chris

-----Original Message-----

Fr m: Davis, Minh-Tam
Sent: Wednesday, May 08, 2002 10:53 AM
To: Chan, Christina
Subject: Rush search request for 09/802520

Please in commercial data base and in issued patent files

1) SEQ ID NO:1 and 2.

Thank you.

MINH TAM DAVIS

ART UNIT 1642, ROOM 8A01, MB 8E12

305-2008

CRFE

Edward Hart
Technical Info. Specialist
STIC/Biotech
CMI 6B02 Tel: 305-9203

FOR OFFICIAL USE ONLY

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 5/8/02
Date Completed: 5/2/02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: 1
AA Sequences: 1
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST(where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: OS
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

Thu May 9 08:44:29 2002

us-09-802-520-1.ra1

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2002, 00:03:01 ; Search time 31.89 Seconds
(without alignments)
375.307 Million cell updates/sec

Title: US-09-802-520-1

Perfect score: 2534
Sequence: 1 MSISMGSPKSLSECLPN.....LEEGIGGTHVSPERYVM 490

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCtus.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	901	35.6	173	US-09-323-873A-8
2	788	31.1	339	US-09-323-873A-2
3	679	26.8	141	US-09-083-521-1
4	144.5	5.7	227	US-09-655-270A-15
5	144.5	5.7	227	US-09-651-941-17
6	118.5	4.7	531	US-08-724-974A-2
7	112.5	4.4	365	US-08-724-974A-3
8	107	4.2	34	US-09-323-873A-20
9	106	4.2	695	US-08-487-886-2
10	106	4.2	695	US-08-482-855-2
11	101	4.0	476	US-09-316-083-3
12	97.5	3.8	390	US-08-460-576-2
13	93	3.7	692	US-07-757-342D-6
14	89.5	3.5	495	US-07-841-997A-2
15	89.5	3.5	495	US-08-290-301-2
16	89.5	3.5	495	US-08-588-983-2
17	89.5	3.5	495	US-08-588-976-2
18	89.5	3.5	861	US-09-022-875-2
19	89.5	3.5	968	US-08-651-999A-7
20	89.5	3.5	968	US-09-385-752-7
21	89	3.5	467	US-08-805-118-3
22	89	3.5	467	US-09-391-958-3
23	89	3.5	1582	US-08-404-531B-9
24	89	3.5	1582	US-08-476-900A-9
25	89	3.5	1582	US-08-488-546A-9
26	89	3.5	1582	US-08-726-320-5
27	89	3.5	1582	US-09-208-716-5

28	88.5	3.5	591	US-08-484-840-2	Sequence 2, Appl1
29	88.5	3.5	591	US-08-483-094-2	Sequence 2, Appl1
30	88.5	3.5	699	US-08-348-006A-7	Sequence 7, Appl1
31	88.5	3.5	699	US-08-800-825A-7	Sequence 7, Appl1
32	88.5	3.5	699	US-09-158-657-7	Sequence 7, Appl1
33	88	3.5	920	US-08-930-996A-8	Sequence 8, Appl1
34	86	3.4	205	US-08-839-711-1	Sequence 1, Appl1
35	86	3.4	480	US-08-724-394A-9	Sequence 9, Appl1
36	85	3.4	337	US-08-153-848-46	Sequence 46, Appl1
37	85	3.4	337	US-09-299-843A-46	Sequence 46, Appl1
38	85	3.4	337	US-09-088-337B-46	Sequence 46, Appl1
39	85	3.4	337	PCT-US93-11153-46	Sequence 46, Appl1
40	85	3.4	352	US-08-202-056-3	Sequence 3, Appl1
41	85	3.4	352	US-08-076-093A-4	Sequence 4, Appl1
42	85	3.4	352	US-08-701-265-4	Sequence 4, Appl1
43	85	3.4	352	US-08-284-586-4	Sequence 4, Appl1
44	85	3.4	352	US-08-805-478-4	Sequence 4, Appl1
45	85	3.4	352	US-08-802-627A-4	Sequence 4, Appl1

ALIGNMENTS

```
RESULT 1
US-09-323-873A-8
; Sequence: 8, Application US/09323873A
; Patent No. 6329503
; GENERAL INFORMATION:
; APPLICANT: Daniel E. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Kahen Leong
; APPLICANT: Arthur B. Raitano
; APPLICANT: Douglas C. Saffran
; APPLICANT: Steve Chappell Mitchell
; TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
; TITLE OF INVENTION: EXPRESSED IN HUMAN CANCERS AND USES THEREOF
; FILE REFERENCE: 129.160502
; CURRENT APPLICATION NUMBER: US/09/323.873A
; CURRENT FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/087,520
; PRIOR FILING DATE: 1998-06-01
; PRIOR APPLICATION NUMBER: 60/091,183
; PRIOR FILING DATE: 1998-06-30
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-323-873A-8

Query Match 35.6%; Score 901; DB 4; Length 173;
Best Local Similarly 100.0%; Pred. No. 1.8e-85;
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 DEYKPIEIVNKTPIYAIPLISVYLAGLAAVLYGKRYKRPFWLETWLOCKROL 305
DB 1 DEYKPIEIVNKTPIYAIPLISVYLAGLAAVLYGKRYKRPFWLETWLOCKROL 60
QY 306 GLISFFPAMVAVASLCLPMRSERYFLNNAVOQVANIENSNNEEVRRIEYISFGI 365
DB 61 GLISFFPAMVAVASLCLPMRSERYFLNNAVOQVANIENSNNEEVRRIEYISFGI 120
QY 366 MSGLGLSLAVTSIPYSNALNMRSEFIOSTGLGYVALLSTFHVLIYGMKRA 418
DB 121 MSGLGLSLAVTSIPYSNALNMRSEFIOSTGLGYVALLSTFHVLIYGMKRA 173

RESULT 2
US-09-323-873A-2
; Sequence 2, Application US/09323873A
; Patent No. 6329503
```

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

AUTHORS Gattung, S. and Maggi, L.
 TITLE The sequence of H. sapiens BAC clone RG016J04
 JOURNAL Unpublished (1997)
 REFERENCE 2 (bases 1 to 156214)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (09-MAY-1997)
 COMMENT Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63108, USA
 http://genome.wustl.edu/gsc
 e-mail: sapientewatson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see http://www.nhgri.nih.gov/DIR/STB/CHR7 or send an E-mail to egreen@nhgri.nih.gov

SOURCE INFORMATION:
 This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci. 89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).
 VECTOR: pBELO
 Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
 Actual start of this clone is at base position 1 of H_RG016J04; actual end is at 156214 of H_RG016J04. The orientation of this clone is unknown.

This clone contains STS SWSS2784 (NTD:g1113580) and SWSS893 (NTD:g454733).

FEATURES

source

Location/Qualifiers

1. 156214

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/map="7q21"

/clone="RG016J04"

/clone.lib="CIRB-978SK-B"

15. 40

/rpt_family="L1"

326. 621

/rpt_family="ALU"

complement(977. 1499)

/rpt_family="L1"

3398. 3421

/rpt_family="L1"

5319. 5345

/rpt_family="L1"

complement(8071. 8347)

/rpt_family="ALU"

complement(9406. 9975)

/rpt_family="L1"

complement(10000. 11285)

/rpt_family="L1"

complement(11315. 11984)

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

THIS PAGE BLANK (USPTO)

```

repeat_region      /rpt_family="PTR"
39839 .40125
repeat_region      /rpt_family="ALU"
40126 .40157
repeat_region      /rpt_family="L1"
complement(41340) .41706
repeat_region      /rpt_family="ALU"
complement(41904) .41933
repeat_region      /rpt_family="L1"
42348 .42409
repeat_region      /rpt_family="ALU"
complement(42942) .42982
repeat_region      /rpt_family="L1"
complement(44380) .44423
repeat_region      /rpt_family="L1"
44659 .44954
repeat_region      /rpt_family="ALU"
45422 .45448
repeat_region      /rpt_family="L1"
complement(45615) .45669
repeat_region      /rpt_family="L1"
complement(45672) .45963
repeat_region      /rpt_family="ALU"
complement(45964) .46597
repeat_region      /rpt_family="L1"
46100 .46516
repeat_region      /rpt_family="L1"
48370 .48785
repeat_region      /rpt_family="L1"
complement(49577) .50216
repeat_region      /rpt_family="L1"
complement(50540) .50791
repeat_region      /rpt_family="TBE"
complement(50820) .50985
repeat_region      /rpt_family="TBE"
complement(51124) .51156
repeat_region      /rpt_family="L1"
51405 .51691
repeat_region      /rpt_family="ALU"
complement(51721) .51764
repeat_region      /rpt_family="L1"
complement(51765) .52058
repeat_region      /rpt_family="ALU"
complement(52252) .52467
Query Match      28.2% Score 533.4; DB 9; Length 156214;
Best Local Similarity 99.8%; Pred. No. 2e-124;
Matches 534; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 804 caggttatatagagaacataatccaagcgagacaagattgaactggccgc 863
|||||
Db 19241 CAGGTTTATATATGAGCAACAAATTCAACGCGACAAAGTTTGAACCTGCCGC
OY 864 cagttgaattcattccattgagcttgagctcattatcatcagcagagagattgaaat 923
|||||
Db 19301 CAGTTGAATTCATTTCCATTGAGCTTGAGGATCCATTATCATCAGCCAGAGATTGAAAT
OY 924 ttaccctcagagctcttactctctgagagagggcagctgtgagctataaagctggcc 983
|||||
Db 19361 TTACCCCTAGAGACTCTTACTCTGTGAGAGGCGCACTGGGTAGCTATTAAGCTTGGCC
OY 984 acatttttttcttattctctctgtcagagatgtatcatcatcatagctagaaccaa 1043
|||||
Db 19421 ACAATTTTCTTCTTATCTCTTGTGAGAGATGTATTCATTCATATAGCTTAACAAACAA
OY 1044 cagagtgacttttacaacattcctatagagattgtgaaataaacccttaactagtgc 1103
|||||
Db 19481 CAGAGTGACTTTTACAAATTCCTATATAGAGATTGGAATAAACCCTTACTTATAGTGC
OY 1104 attacttctctcctcctagatatacctcagaggtctctctgagcgtcgtatcaacttat 1163
|||||
Db 19541 ATTACTTGTCTCTCCTCTAGTATACCTCGCAGAGTCTTCTGCGAGCTGCTTATCACTTAT 19600

```

```

OY 1164 taggcacccaagtatagagattccaccctgtgttggaaccttggtacagtagaanaa 1223
|||||
Db 19601 TAGGCAACCAAGTATAGAGATTTCACACCTTGTTGGTAACCTGGTTACAGTGTAGAAA 19660
OY 1224 cagcttgattactaagttttttcttcgtctatggtccatgltgcctacagcctctgcta 1283
|||||
Db 19661 CACTTGGATTACTAAGTTTTTCTTCGTATGCTATGCTCATCTTCCCTACACCTCTGCTTA 19720
OY 1284 ccgattagagaagtcagagagatatgttttcttaacatggtctatacagcagctc 1338
|||||
Db 19721 CCGATGAGAGAGTCACAGAGATATTGTGTTTCACACATGCTTATCAGCAGGTAC 19775

RESULT 3
AC104475/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Ho,S.-L., Ido,J.R., Karlins,E., Laric,P., Lee-Tsin,S.-O.,
Legaspi,R., Maduro,O.L., Maduro,V.B., Mastello,C., Mastrian,S.D.,
McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A., Stantillo,S.,
Thomas,M.A., Thomas,P.J., Touchman,J.W., Tsurgou,C., Vogt,J.L.,
Walker,M.A., Wetherby,K.D., Whigins,L., Young,A., Zhang,L.-H. and
Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 162928)
Direct Submission
Submitted (12-DEC-2001) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gaithersburg, MD 20877, USA

----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@nih.gov
----- Project Information
Center project name: ces
Center clone name: 120K11

----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 161001 bases at least Q40
Consensus quality: 161422 bases at least Q30
Consensus quality: 161562 bases at least Q20
Insert size: 14300; agarose-1p
Insert size: 162428; sum-of-ctnigs
Quality coverage: 10.76x in Q20 bases; agarose-1p
Quality coverage: 9.47x in Q20 bases; sum-of-ctnigs

```

NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 13922: contig of 13922 bp in length
 * 13923 14022: gap of unknown length
 * 14023

THIS PAGE BLANK (USPTO)

```

/Note="unnamed protein product"
/codon_start=1
/protein_id="CAC42678.1"
/db_xref="GI:14536714"
/translation="MESISMSPKSLSETCLPGLNGINGIDKARYVYVYGSDFPKS
LIRLIRCCGHVIGSRNPKFSEFPFPHVVDVHHEDALTKTNIIVVAIHREHYTSLM
ASROYVICSNNIOAROOVIELAROLNFIPIDIGSLSSAREIENLPLRFTLMGPVYV
AASLAFYFLYSFVRDVIHPVARNOSDYPKIPLEIVNKTLPVATVLTSLVYLGGL
MAAYOLYGTGRFPPELTPVNLQCKKQGLSFFAFMVAHVAVAVSICLPRRSERYLFL
FIOSLGVANINENSNBEKRWIRIEMVIFSGIMSLGLSLIATVSTIFSSNALMWEFS
RHPD"
BASE COUNT      648 a      537 c      520 g      748 t
ORIGIN

```

```

Query Match      44.3%. Score 229; DB 6; Length 2453;
Best Local Similarity 100.0%; Pred. No. 2.3e-128;
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 280 aggatattcttgatgattggaagtcgcgtatcatcaggaaatcctcctatgagaa 339
Db 320 AGGATATTCTTGTGATCTTGAGAGTCTCGATATCATGATCAATCTCTATGATGGAA 379
Qy 340 gccctaaagccttaagtgaaactgtttacctaataatgataatgataaagatgcaa 399
Db 380 GCCCTAAGACCTTGTAGTGAACCTGTTACTTATGATCAATGATGATCAAGATGCAA 439
Qy 400 ggaagtcactgtagtgatggaatggaatggaatggaatggaatggaatggaatggaat 459
Db 440 GGAGGTCACCTGATGATGATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 499
Qy 460 ttattagaagcggtatcatcgtgcatgaggaatggaatggaatggaatggaatggaatggaat 508
Db 500 TTATTAGATGCGGTATCATGTCGTATGATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 548

```

```

RESULT 2
LOCUS      HSAC002064      156214 bp      DNA      linear      PRI 09-MAY-1997
DEFINITION Human BAC clone RG016J04 from 7q21, complete sequence.
ACCESSION AC002064
VERSION    AC002064.1 GI:2076723
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1 (bases 1 to 156214)
            Gattung S. and Magg1, L.
            The sequence of H. sapiens BAC clone RG016J04
            Unpublished (1997)
AUTHORS    Waterston, R.
            Direct Submission
            Submitted (09-MAY-1997)
            Genome Sequencing Center
            Department of Genetics, Washington University
            St. Louis, MO 63108, USA
            http://genome.wustl.edu/gsc
            e-mail: sapient@watson.wustl.edu
JOURNAL
TITLE
COMMENT

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was

confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NIGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTH/CHR7> or send an E-mail to egreen@nhgri.nih.gov

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 89.8794-8797 (1992); Kim et al., Proc. Natl. Acad. Sci. VECTOR: pBelo Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of H_RG016J04; clone end is at 156214 of H_RG016J04. The orientation of this clone is unknown.

This clone contains STS SWS52784 (NID: g1113580) and SWS5893 (NID: g454733).

FEATURES	source
repeat_region	1. 156214
repeat_region	/organism="Homo sapiens"
repeat_region	/db_xref="taxon:9606"
repeat_region	/chromosome="7"
repeat_region	/map="7q21"
repeat_region	/clone="RG016J04"
repeat_region	/clone.lib="CITB-978SK-B"
repeat_region	15. 40
repeat_region	/rpt_family="L1"
repeat_region	326. 621
repeat_region	/rpt_family="ALU"
repeat_region	complement(977. 1499)
repeat_region	/rpt_family="L1"
repeat_region	3398. 3421
repeat_region	/rpt_family="L1"
repeat_region	5319. 5345
repeat_region	/rpt_family="L1"
repeat_region	complement(8071. 8347)
repeat_region	/rpt_family="ALU"
repeat_region	complement(9406. 9975)
repeat_region	/rpt_family="L1"
repeat_region	complement(10000. 11285)
repeat_region	/rpt_family="L1"
repeat_region	complement(11315. 11984)
repeat_region	/rpt_family="L1"
repeat_region	11666. 11687
repeat_region	/rpt_family="L1"
repeat_region	complement(12010. 12299)
repeat_region	/rpt_family="ALU"
repeat_region	complement(12301. 13893)
repeat_region	/rpt_family="L1"
repeat_region	13897. 14096
repeat_region	/rpt_family="L1"
repeat_region	complement(14777. 14838)
repeat_region	/rpt_family="L1"
repeat_region	15715. 15767
repeat_region	/note="similar to human EST T02878 (NID: g314119)."
repeat_region	19436. 19497
repeat_region	/note="similar to human EST AA123941 (NID: g1682616)"
repeat_region	mq22f09.r1
repeat_region	complement(25113. 25148)
repeat_region	/rpt_family="L1"
repeat_region	complement(25561. 25578)
repeat_region	/rpt_family="L1"
repeat_region	complement(26832. 27124)
repeat_region	/rpt_family="ALU"
repeat_region	27125. 27383

THIS PAGE BLANK (USPTO)

Query Match	Best Local Similarity	34.4%	Score 178:	DB 9:	Length 156214:
Matches 228:	Conservative	99.6%	Pred. No. 7.5e-97:	Mismatches 1:	Indels 0:
				Gaps 0:	
repeatL_region	/rptL_family="L1"	complement(45964 .46597)			
repeatL_region	/rptL_family="L1"	46100 .46516			
repeatL_region	/rptL_family="L1"	48370 .48785			
repeatL_region	/rptL_family="L1"	complement(49577 .50216)			
repeatL_region	/rptL_family="L1"	complement(50540 .50791)			
repeatL_region	/rptL_family="RHE"	complement(50820 .50985)			
repeatL_region	/rptL_family="RHE"	complement(51124 .51156)			
repeatL_region	/rptL_family="L1"	51405 .51691			
repeatL_region	/rptL_family="ALU"	complement(51721 .51764)			
repeatL_region	/rptL_family="L1"	complement(51765 .52058)			
repeatL_region	/rptL_family="ALU"	complement(52252 .52467)			
repeatL_region	complement(52252 .52467)				
Db 280	aggaatactcttgatccttggaagtgctgcgtacatacgaatcaatcctatgatgggaa	339			
Db 17321	AGGATATTCCTTGATCTTGGAGAGTGCCTGATCATGATCAATCATCTATGATGGGAA	17380			
Db 340	ggccctaaagagccttgatgaacttgcttcccaagagatgaatgatcaagaatggaa	399			
Db 17381	GGCCCTAAGAGCCTTGATGTAACCTTTTTCATGATGCAATATGATGATCAAGATGCAA	17440			
Db 400	ggaagtgctacgtgaatggaatggaagtggaagatttcgaatccctgaccattggac	459			
Db 17441	GGAGAGTCACTCTAGCTGATTCGATTCGAAGTGGAGATTTTCCCAATCCTTGACATTGGAC	17500			
Db 460	ttaataagatgcgcatacatgtgtcatatggaagtagaatactctaagt	508			
Db 17501	TTATATAGATGCGGCTATCATGTGTCATGAGAAATATCTTAAGTT	17549			
RESULT 3	AC104475	162928 bp	DNA	linear	HTG 12-DEC-2001
AC104475/C	AC104475	162928 bp	DNA	linear	HTG 12-DEC-2001
LOCUS	pan troglodytes clone RP43-120K11, WORKING DRAFT SEQUENCE, 6				
DEFINITION	unprocessed pieces.				
ACCESSION	AC104475				
VERSION	AC104475.1	GI:17530717			
KEYWORDS	HTG: HTGS-PHASE1; HTGS-DRAFT.				
SOURCE	Chimpanzee.				
ORGANISM	Pan troglodytes				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.				
	1 (bases 1 to 162928)				
	Akter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,				
	Blakesley,R.V., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,				
	Dietrich,N.L., Granter,S., Guan,X., Gupta,J., Haghighi,P.,				
	Ho,S.-L., Idol,J.R., Karlins,E., Latic,P., Lee-Il-in,S.-O.,				
	Legaspi,R., Maduro,O.L., Maduro,V.B., Mastello,C., Mastrian,S.D.,				
	McMloskey,J.C., McDowell,J., Pearson,R., Prasad,A., Stantirpop,S.,				
	Thomas,J.W., Thomas,P.J., Touchman,J.W., Tsurgon,C., Vogt,J.L.,				
	Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and				
	Green,E.D.				
	NISC Comparative Sequencing Initiative				
	Unpublished				
	2 (bases 1 to 162928)				
TITLE	Direct Submission				
REFERENCE	Green,E.D.				
AUTHORS	Submitted (12-DEC-2001) NIH Intramural Sequencing Center, 8717				
JOURNAL	Grovenom Circle, Galtersburg, MD 20877, USA				
COMMENT	----- Genome Center				

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2002, 21:01:15 ; Search time 3502.82 Seconds
(without alignments)
11297.206 Million cell updates/sec

Title: US-09-802-520-2

Perfect score: 1891

Sequence: 1 ggggaagcagctgagtgctg.....gtcaattatcgtgggttga 1891

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
------------	-------	-------	-------	--------	----	----	-------------

1	1409.8	74.6	2453	6	AX155249	AX155249 Sequence
2	533.4	28.2	156214	6	HSAC002064	AC002064 Human BAC
3	528.6	28.0	162828	2	AC1004475	AC1004475 Pan trogl
4	516.2	27.3	172915	2	AC099742	AC099742 Papio cyn
5	425.8	22.5	444	6	AX106433	AX106433 Sequence
6	425.8	22.5	444	6	AX140724	AX140724 Sequence
7	425.8	22.5	444	6	AX200584	AX200584 Sequence
8	425.8	22.5	444	6	AX267240	AX267240 Sequence
9	415.2	22.0	1971	10	AY029586	AY029586 Mus muscu
10	411	21.7	205085	2	AC026813	AC026813 Mus muscu
11	411	21.7	227144	2	AC092404	AC092404 Mus muscu
12	406.6	21.5	2707	10	AF335281	AF335281 Rattus no
13	381	20.1	3844	6	AX211372	AX211372 Sequence
14	379.4	20.1	1845	9	AY029585	AY029585 Homo sapi
15	379.4	20.1	3912	9	AK001691	AK001691 Homo sapi
16	363	19.2	366	6	AX106434	AX106434 Sequence
17	363	19.2	366	6	AX140725	AX140725 Sequence
18	363	19.2	366	6	AX200585	AX200585 Sequence
19	363	19.2	366	6	AX267241	AX267241 Sequence
20	358.2	18.9	1858	6	AX083516	AX083516 Sequence
21	356.6	18.9	4429	6	AX155251	AX155251 Sequence
22	337.8	17.9	2462	2	AK026806	AK026806 Homo sapi
23	335.2	17.7	87401	2	AC021898	AC021898 Homo sapi
24	318.2	16.8	3155	10	BC006651	BC006651 Mus muscu
25	317.8	16.8	1763	10	AY029778	AY029778 Mus muscu
26	315	16.7	3080	10	MMU0319746	MMU0319746 Mus muscu
27	310.2	16.4	322	6	AX155253	AX155253 Sequence
28	306.8	16.2	328	6	AX106431	AX106431 Sequence
29	306.8	16.2	328	6	AX140722	AX140722 Sequence
30	306.8	16.2	328	6	AX200582	AX200582 Sequence
31	306.8	16.2	328	6	AX267238	AX267238 Sequence
32	301.6	15.9	1754	4	AF319659	AF319659 Sus scrofa
33	290.4	15.4	1193	6	AX155243	AX155243 Sequence
34	290.4	15.4	1195	6	AX201105	AX201105 Sequence
35	290.4	15.4	1195	6	AX267904	AX267904 Sequence
36	290.4	15.4	1195	9	AF186249	AF186249 Homo sapi
37	290.4	15.4	1249	9	BC011802	BC011802 Homo sapi
38	272.8	14.4	1175	10	AF297098	AF297098 Homo sapi
39	266.4	14.1	1191	10	AY029584	AY029584 Mus muscu
40	243.4	12.9	250	6	AX106432	AX106432 Sequence
41	243.4	12.9	250	6	AX140723	AX140723 Sequence
42	243.4	12.9	250	6	AX200583	AX200583 Sequence
43	243.4	12.9	250	6	AX267239	AX267239 Sequence
44	232.2	12.3	3627	6	AX155248	AX155248 Sequence
45	228.4	12.1	1203	6	A63553	A63553 Sequence 6

ALIGNMENTS

RESULT 1
AX155249
LOCUS AX155249 2453 bp
DEFINITION Sequence 7 from Patent WO0140276.
ACCESSION AX155249
VERSION AX155249.1 GI:14536713
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
1 (bases 1 to 2453)
Farris, M. and Jakobovits, A.
Serpentine transmembrane antigens expressed in human prostate
cancers and uses thereof
Patent: WO 0140276-A 7 07-JUN-2001;
JOURNAL Urogenesys, Inc. (US)
FEATURES
Location/Qualifiers
source
1..2453
/organism="Homo sapiens"
/db_xref="taxon:9606"
355..1719

AUTHORS	Galling, S. and Maggi, L.
TITLE	The sequence of H. sapiens BAC clone RG016J04
JOURNAL	Unpublished (1997)
REFERENCE	2 (bases 1 to 156214)
AUTHORS	Waterston, R.
TITLE	Direct Submission
JOURNAL	Submitted (09-MAY-1997)
COMMENT	Genome Sequencing Center

Department of Genetics, Washington University
St. Louis, MO 63108, USA
<http://genome.wustl.edu/gsc>
e-mail: sapiens@watsn.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7> or send an E-mail to sgreen@nhgri.nih.gov

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. For references see: Shizuya et al., *Proc. Natl. Acad. Sci.*, 89:8794-8797 (1992); Kim et al., *Genomics* 34:213-218 (1996).

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of H_RG016J04; actual end is at 156214 of H_RG016J04. The orientation of this clone is unknown.

This clone contains STS SWSS2784 (NID:g1113580) and SWSS8933 (NID:g454733).

FEATURES

Source

```

repeat_region
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="7"
  /map="7q21"
  /clone="K0016704"
  /clone_id="CTB-978SK-B"
  15..40
  /rpt_family="L1"
  326..621
  /rpt_family="ALU"
  complement(977..1499)
repeat_region
  /rpt_family="L1"
  3398..3441
  /rpt_family="L1"
  5319..5345
  /rpt_family="L1"
  complement(8071..8347)
  /rpt_family="ALU"
  complement(9406..9975)
  /rpt_family="L1"
  complement(11000..11285)
  /rpt_family="L1"
  complement(11315..11984)
repeat_region

```

repeat_region	/rpt_family="L1" 11666. .11687 /rpt_family="L1" complement(12010. /rpt_family="ALU" complement(12301. /rpt_family="L1" 13897. .14096 /rpt_family="L1" complement(14777. /rpt_family="L1" 15715. .15767 /note="similar to human EST T02878 (NID:g314119)" /note="similar to human EST AA123941 (NID:g1682616) mq22f09.r1"
repeat_region	complement(25113. /rpt_family="L1" complement(25561. /rpt_family="L1" complement(26832. /rpt_family="ALU" 27125. .27383 /note="similar to human EST AA151807 (NID:g1720502) z139g02.r1"
repeat_region	27125. .27544 /note="similar to human EST AA151796 (NID:g1720491) z139c02.r1"
repeat_region	complement(27178. .27600) /note="similar to human EST AA149579 (NID:g1720380) z139c02.s1"
misc_feature	complement(27441. .27890) /note="similar to human EST N52554 (NID:g1193720) yv36a11.s1"
misc_feature	27686. .27856 /note="similar to human EST W22120 (NID:g1313113) zb97c08.r1"
misc_feature	27686. .27838 /note="similar to human EST W32085 (NID:g1313105) zb97b08.r1"
misc_feature	complement(27733. .28143) /note="similar to human EST N59831 (NID:g1203721) yz27a10.s1"
misc_feature	complement(27838. .28143) /note="similar to human EST W31561 (NID:g1312680) zb97b08.s1"
misc_feature	complement(27887. .28143) /note="similar to human EST N59830 (NID:g1203720) yz27a09.s1"
misc_feature	complement(27997. .28144) /note="similar to human EST W31628 (NID:g1312688) zb97c08.s1"
misc_feature	28666. .28972 /note="similar to human EST R09339 (NID:g761262) yf26d05.r1"
misc_feature	complement(29587. .29853) /note="similar to human EST R09227 (NID:g761150) yf26d05.s1"
misc_feature	30146. .30314 /note="similar to human EST N78042 (NID:g1240743) yv1e06.r1"
misc_feature	complement(31170. .31350) /note="similar to human EST N58451 (NID:g1202341) yv1e06.s1"
repeat_region	31970. .32006 /rpt_family="L1"
repeat_region	32263. .32563 /rpt_family="ALU"
repeat_region	complement(34737. .34778) /rpt_family="L1"
repeat_region	36558. .36709 /rpt_family="ALU"
repeat_region	complement(37630. .37740)

```

repeat_region      /rpl_family="PTR"
                    39839. .40125
                    /rpl_family="ALU"
repeat_region      40126. .40157
                    /rpl_family="L1"
repeat_region      complement(41340. .41706)
                    /rpl_family="ALU"
repeat_region      complement(41904. .41933)
                    /rpl_family="L1"
repeat_region      42348. .42409
                    /rpl_family="ALU"
repeat_region      complement(42942. .42982)
                    /rpl_family="L1"
repeat_region      complement(44380. .44423)
                    /rpl_family="L1"
repeat_region      44659. .44954
                    /rpl_family="ALU"
repeat_region      45422. .45448
                    /rpl_family="L1"
repeat_region      complement(45615. .45669)
                    /rpl_family="L1"
repeat_region      complement(45672. .45963)
                    /rpl_family="ALU"
repeat_region      complement(45964. .46397)
                    /rpl_family="L1"
repeat_region      46100. .46516
                    /rpl_family="L1"
repeat_region      48370. .48785
                    /rpl_family="L1"
repeat_region      complement(49577. .50216)
                    /rpl_family="L1"
repeat_region      complement(50540. .50791)
                    /rpl_family="TBE"
repeat_region      complement(50820. .50985)
                    /rpl_family="TBE"
repeat_region      complement(51124. .51156)
                    /rpl_family="L1"
repeat_region      51405. .51691
                    /rpl_family="ALU"
repeat_region      complement(51721. .51764)
                    /rpl_family="L1"
repeat_region      complement(51765. .52058)
                    /rpl_family="ALU"
repeat_region      complement(52252. .52467)

```

Query Match 28.2%; Score 533.4; DB 9; Length 156214;
 Best Local Similarity 99.8%; Pred. No. 2e-124;
 Matches 534; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 804 cagggttatataatgacgaacaatatcaagcgacacacaggttatgaactgcccgc 863
DB 19241 CAGGTTATATATGACGACCAATATTCAGCCGACACAGGTTATGAACTTGCCCGC
QY 864 cagttgaattcctccatctggtggtggtccttatcaccacagagattgaaat 923
DB 19301 CAGTTGATTTATCTCCATGACCTGGGATCCCTTATCATCAGCCAGAGATTGAATAAT
QY 924 ttcccccctacgactcttactctctggaagagggcaggtggtgtaactaagctggcc 983
DB 19361 TTAACCCCTAGACCTTTACTCTCTGAGAGAGGCCAGAGTGTGAGCTATAAGCTTGCC
QY 984 acatttttttcttactctctgtcagagagatgattcatcatatgctagaacaa 1043
DB 19421 ACATTTTTCCTTATCTTTGTCAGAGATGATTCATTCATGCTAGAAACAA
QY 1044 cagaagacactttacaaaatctcctatagatgtaataaaactactatagttgcc 1103
DB 19481 CAGAGGACTTTTACAAAATCTCATAGAGATTGTGAATAAACCTTATAGTTGCC
QY 1104 attacttctctcctcctagatatactcgcagctctctcgcagctgcttaacattat 1163
DB 19541 ATTACTTTGCTCTCCCTAGATATACCTGCGAGGCTTCTCTGACACTGCTTATCAACTTAT 19600

```

```

QY 1164 tacggcaccaagatagagattccacctggttggaacccggtgtacagtgtaagaaa 1223
DB 19601 TACGGCACCAAGATATAGAGATTTCACCTTGTTGGAACCTGCTTACAGTGAGAAAA 19660
QY 1224 cagcttgatctactaaattttctctgctatagttcattgttcctacagcctctgctta 1283
DB 19661 CAGCTTGATTTACTAAATTTTTCTTCGCTATGCTCATGTCCTTACACCTCTGCTTA 19720
QY 1284 ccgattggaaggtcagaagatatctgtctcaacatgcttatcagcaggttc 1338
DB 19721 CCGATTGAGAGGTCTAGAGATATTTCTTCTCATCATGCTTATACAGAGTAC 19775

```

RESULT 3
 AC104475/c
 LOCUS
 DEFINITION
 pan troglodytes clone RP43-120K11, WORKING DRAFT SEQUENCE, 6
 UNORDERED PIECES.
 AC104475 162928 bp DNA linear HTG 12-DEC-2001
 AC104475.1 GI:17530717
 VERSION
 AC104475.1 GI:17530717
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS
 chimpanzee.
 SOURCE
 Pan troglodytes
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
 1 (bases 1 to 162928)
 Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
 Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
 Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
 Ho,S.-L., Idol,J.R., Karlins,E., Lait,P., Lee,Lin,S.-Q.,
 Legaspi,R., Maduro,Q.L., Maduro,V.B., Masello,C., Mastrian,S.D.,
 McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A., Stantipop,S.,
 Thomas,J.W., Thomas,P.J., Touchman,J.W., Tsugeon,C., Vogt,J.L.,
 Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and
 Green,E.D.

TITLE
 JOURNAL
 NISC Comparative Sequencing Initiative
 REFERENCE
 2 (bases 1 to 162928)
 Green,E.D.
 AUTHORS
 Direct Submission
 JOURNAL
 Submitted (12-DEC-2001) NIH Intramural Sequencing Center, 8717
 Grovemont Circle, Gaithersburg, MD 20877, USA
 COMMENT
 Genom Center
 Center: NIH Intramural Sequencing Center
 Center code: NISC
 Web site: http://www.nisc.nih.gov
 Contact: nisc.mouse@nih.gov
 Project Information
 Center project name: ccs
 Center clone name: 120K11
 Summary Statistics
 Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 161001 bases at least Q40
 Consensus quality: 161422 bases at least Q30
 Consensus quality: 161562 bases at least Q20
 Insert size: 143000; agarose-ef
 Insert size: 162428; sum-of-ctgts
 Quality coverage: 10.76x in Q20 bases; agarose-ef
 Quality coverage: 9.47x in Q20 bases; sum-of-ctgts

NOTE: This is a 'working draft' sequence. It currently
 consists of 6 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.
 1 13922: contig of 13922 bp in length
 * 13923 14022: gap of unknown length
 * 14023 36248: contig of 22226 bp in length

```

*      36249      36348: gap of unknown length
*      36349      66192: contig of 29844 bp in length
*      66193      66292: gap of unknown length
*      66293      92168: contig of 25876 bp in length
*      92169      92268: gap of unknown length
*      92269      122036: contig of 29768 bp in length
*      122037      122137: gap of unknown length
*      122137      162928: contig of 40792 bp in length.
FEATURES
SOURCE
1. 162928
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="RP43-120K11"
/clone_1ib="RP43"
1. 13922
/note="assembly_fragment"
clone_end:17
vector_side:left"
14023. 36248
/note="assembly_fragment"
36349. 66192
/note="assembly_fragment"
66293. 92168
/note="assembly_fragment"
92269. 122036
/note="assembly_fragment"
122137. 162928
/note="assembly_fragment"
clone_end:SP6
vector_side:right"

```

```

BASE COUNT      51528 a 29878 c 30533 g 50485 t 504 others
ORIGIN

```

```

Query Match      28.0%, Score 528.6; DB 2; Length 162928;
Best Local Similarity 99.3%; Pred. No. 3.4e-123;
Matches 531; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      804      caggtttatatgagaacaatttcaagcgagacaacaggttattgaacttccgcg 863
Db 121803 CAGGTTTATATATGAGACACATATTCAGCCCGACAAACAGGTTATGTAACCTTCCCGC 121744

QY      864      cagtgtaattcattcccatctgaacttggaatccatcatcagcagagagaattgaat 923
Db 121743 CAGTTGATTTTCATTCCTCCATTCAGTTGGATCCTTATCATCATCCAGAGAGATTTGAANAAT 121684

QY      924      ttaccctcagactcttactctctcgagagggccagtggtgtgtactataaagcttgcc 983
Db 121683 TTACCCCTACGACCTTACTCTCTGAGAGGGCCAGTGGTGTGTATAGCTTGGCC 121624

QY      984      acatttttttcttattctcttcttgtagagatgtgatccatccatagtctagaaccaa 1043
Db 121623 ACATTTTTCCTTATTCCTTTCCTGTCAGAGATGTGATTCATTCATGTAAGAAACCA 121564

QY      1044      cagaatgactttacaacattccatagagatgtgaaataaacttaacttaacttgcc 1103
Db 121563 CAGAGTGAATTTTACAAAATTCCTTATAGAGATGTGAATTAACCTTACTTATGTTGCC 121504

QY      1104      attacttgcctccctagatctacccgcagagctcttcctcgacagctgctatcaacttat 1163
Db 121503 ATTACTTGTCTCTCTCTAGTATACCTCGCAGAGCTTCTGGCAGCTGCTTATCAACTTAT 121444

QY      1164      tacggcaccaagtataagaatttccacctggtgtggaacctggtttacagtgtagaaa 1223
Db 121443 TACGGCACCAAGTATAGAGATTTCCACCTGTTGGAAACCTGTTACAGTGTGAANA 121384

QY      1224      cagcttgatctactaagtttttcttcctcgatctgctcagctgctgctacagcctcgctta 1283
Db 121383 CAGCTTGATTTACTAAGTTTTTTCTTCCTCTGCTGTGTCATGTTGCTTACAGCCTCTGCTTA 121324

QY      1284      ccgatgagaaggtcagagagatatattgttctcaacatgctcttaccagcaggttc 1338
Db 121323 CCGATGAGAAGGTCCAGAGAGATATTTGTTCTCAACATGGCTTATCACAGGTAC 121269

```

```

RESULT      4
AC099742
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.mouse@nrl.nih.gov
----- Project Information
Center project name: ccy
Center clone name: 167P22
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 171178 bases at least Q40
Consensus quality: 171910 bases at least Q30
Consensus quality: 172236 bases at least Q20
Insert size: 130000; agarose-fp
Insert size: 17615; sum-of-contigs
Quality coverage: 12.58x in Q20 bases; agarose-fp
Quality coverage: 9.47x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2439      2438: contig of 2438 bp in length
*      2439      2538: gap of unknown length
*      2539      8133: contig of 5555 bp in length
*      8134      8233: gap of unknown length
*      8234      40378: contig of 32145 bp in length
*      40379      172915: contig of 132437 bp in length.
*      40479      172915: contig of 132437 bp in length.
FEATURES
SOURCE
1. 172915
/organism="Papio cynocephalus anubis"
/db_xref="taxon:9555"
/clone="RP41-167P22"
/clone_1ib="RP41"

```

```

misc-feature 1..2438
              /note="assembly_fragment"
              clone_end:SP6
              vector_side:right"
misc-feature 2539..8133
              /note="assembly_fragment"
              /db_xref="taxon:9606"
              /note="assembly_fragment"
              clone_end:77
              vector_side:right"
misc-feature 40479..172915
              /note="assembly_fragment"
              /db_xref="taxon:9606"
BASE COUNT 52509 a 31733 c 32277 g 56096 t 300 others
ORIGIN

Query Match 27.3%; Score 516.2; DB 2; Length 172915;
Best Local Similarity 97.6%; Pred. No. 5e-120; Indels 0; Gaps 0;
Matches 524; Conservative 0; Mismatches 13;

QY 280 aggatattctgtgtagcttggagagtgctgatacatgaatcatctatgtggaa 339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 83188 AGGATATTCTTTGGTGTGATCTGGAGAGTGTGATCATGGAATCATCTGATATGGAA 83247

QY 340 gccctaagagccttagtgaactgtttacctaaaggcataaaatggatcaaatgcaa 339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 83248 GCCCTAAGAGCCTTAGTGAACCTTTTACCTAATGGCATAAATGCTATCAAGATGCAA 83307

QY 400 ggaagtagcacttagtggtagtggagagtgagattggccaatccttgaccctgcagc 459
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 83308 GGAAGTAGCAGTGGTAGTGGAGTGGAGATGTTGGCAATCTTCAGCATTCGAC 83367

QY 460 ttattagatgcggctacatctgtgcataggaagtagaaatccctaagtttgccttcgaat 519
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 83368 TTATTAGATGTGGCTATCATGTGGCTATAGGAAGTAGAAATCTGATGTTGCTTGAAT 83427

QY 520 ttttccatcatgtgtagatgtcaactcatcatgaagaagctctcacaaaaacaataaa 579
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 83428 TTTTCCATCATGTGTAGATGTGCACCCATCATGAATGATGCTGCACAAAACAAATATA 83487

QY 580 tatttgctctacacagaagaacatfataccctccttggagagccggagacactctctg 639
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 83488 TATTGTTCTTACATFATAGAACATTTATACCTCCTGTGGGACCTTAAGACATCTGCTGG 83547

QY 640 tgggtaaaaatccctgattgtatgtgaagaataaagagataaaccagtaaccagaaatcca 699
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 83548 TGGGTAAAATCTGATTTGATGTGAGCAATFATACATGAAGATAAACAGTACCCGAAATCCA 83607

QY 700 atgcggaataatttggtctcataatattcccaagattccttgattgtcaagaagatttaattg 759
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 83608 ATGCTGAATAATTGGGCGCATATTATCCAGATTTCTTTGATTTGCAAGAGATTTAATGTTG 83667

QY 760 tctcagcttgagcaactcagtttagagaccataagatccagccggcaggtttatatat 816
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 83668 TCTCAGCTTGCGCAGCTTCACTTAGACCTTAGAGATGCCAGCCGCGAGTATGTAATT 83724

RESULT 5
LOCUS AX106433 444 bp DNA linear PAT 30-APR-2001
DEFINITION Sequence 214 from Patent WO0125272.
ACCESSION AX106433
VERSION AX106433.1 GI:13922112
KEYWORDS
SOURCE
ORGANISM Homo sapiens
            human.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 444)
AUTHORS Xu,J., Skeiky,Y.A., Reed,S.G. and Cheever,M.A.
TITLE Compositions and methods for therapy and diagnosis of prostate
        cancer
        Patent: WO 0125272-A 214 12-APR-2001;
JOURNAL

```

```

FEATURES             CORIAX CORPORATION (US)
            source      Location/Qualifiers
            misc-feature 1..444
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
BASE COUNT 120 a 100 c 78 g 143 t 3 others
ORIGIN

Query Match 22.5%; Score 425.8; DB 6; Length 444;
Best Local Similarity 98.6%; Pred. No. 1.1e-97; Indels 1; Gaps 1;
Matches 438; Conservative 0; Mismatches 5;

QY 688 acccaaatccaatgctggaatttggctcattatcccaagtttcttgatgtcgaag 747
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 ACCCAGAATCCAAATGCTGAATATTTGGCTTCATTTATCCAGATTTCTTATTTGTCAAAG 60

QY 748 gattaaatgtgtcagcttggcacttcaagttagaccataagatgccaagccggcagg 807
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 GATTTAATGTGTCTAGCTTGGGCACTTCAAGTATGAGACCTTAAGATGCCAGCCGCGAG 120

QY 808 ttatataatgcagcaaatatcatcaagcgcaaaaggttatagacttgcggcagct 867
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 121 TTATATATGACACAAATATTCACCGGACACAGGTTATTTGAACCTTGCCGCGAGT 180

QY 868 tgatttcatctccattggaattggatcctcatcatcagcagagagatgaaatttac 927
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 181 TGAATTTATCCCATTTGATGATTTGGATCTTTATCATCAGCCANAGAGATTTGAAATTTAC 240

QY 928 ccctaagactcttactctcttgagagagcgcaatggtgtagctataagcttggcca- 986
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 241 CCTACAGACTTCTTACTCTCTGAGAGAGGCGCAGTGGGTAGCTATFAGCTTGCCGCACAT 300

QY 987 ttttttcttattccttcttgcagagatgtattcatcctatgtctaaacccaacag 1046
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 301 TTTTTCCTTATTTCTTTGTCAGAGATGCGATTCATCATATGCTATANAACCAACAG 360

QY 1047 agtgaactttacaataatccctatagagatgtgtaataaacccttaactagttgcatt 1106
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 361 AGTGACTTTACAAATTTCTTATAGANATTGTGAATTAACCTTACTATAGTTGCCATT 420

QY 1107 acttgcctccctagatacctc 1130
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 421 ACTTGCCTCTCCTAATATATCTC 444

RESULT 6
LOCUS AX140724 444 bp DNA linear PAT 31-MAY-2001
DEFINITION Sequence 214 from Patent WO0134802.
ACCESSION AX140724
VERSION AX140724.1 GI:14280835
KEYWORDS
SOURCE
ORGANISM Homo sapiens
            human.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 444)
AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
            Reed,S.G., Kalos,M.D., Retter,M.W., Stolk,J.A., Day,C.H.,
            Skeiky,Y.A. and Wang,A.
TITLE Compositions and methods for the therapy and diagnosis of prostate
        cancer
        Patent: WO 0134802-A 214 17-MAY-2001;
JOURNAL CORIAX CORPORATION (US)
            Location/Qualifiers
FEATURES             1..444
            source      /organism="Homo sapiens"
                        /db_xref="taxon:9606"
BASE COUNT 120 a 100 c 78 g 143 t 3 others
ORIGIN

```


Query Match 22.5%; Score 425.8; DB 6; Length 444;
 Best Local Similarity 98.6%; Pred. No. 1.1e-97;
 Matches 438; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 688 accagaatccaatgctgaataattggtctcattatcccaagattcttgatgtcaag 747
 DB 1 ACCCAAAATCCAAATGCTGAATATTGGCTTCATTATCCAGATCTTGATGTCAAG 60

QY 748 gatttaattgtctcagcttgaggcactcaagtgagacctaaagatgcagccgag 807
 DB 61 GATTTAATGTGTGCTCAGCTTGCGCACTTCAGTTAGACCTAAGATGCCAGCCGAG 120

QY 808 ttataatgcagcaacaatattcagcgacacaggttattgacttgcccgagc 867
 DB 121 TTTATATATGCAACAATATTCAAGCGGACACAGATTATTTGAATGCGCCAGT 180

QY 868 tgaattcatcccatctgacttgagatccctatcatcagccagagagattgaattac 927
 DB 181 TGAATTTCAATTCCTCATTTGAGATCCTTATCATCAGCCANAGATTTGAATTTTAC 240

QY 928 cccatgcactcttactctctcagagagcgagtggtgactataagcttgccaca- 966
 DB 241 CCTACGACTCTTACTCTCTGAGAGGCGCAAGTGTGTAAGCTTAAGCTTGCCACAT 300

QY 987 ttttttcttattcctcttgtagagatgtagatccatctgtagaaccacag 1046
 DB 301 TTTTTCCTTTATTCCTTTGTCAGAGATGCAATTCATTCCTANAAACCAACAG 360

QY 1047 agtgaatttacaatctcctataagagattgtaataaacttactactagttgccatt 1106
 DB 361 AGTGACTTTTACAAATTCCTATAGANATTGTGAATTAACCTTACTTATGTCGCAAT 420

QY 1107 acttgctctccctagatactc 1130
 DB 421 ACTTGTCTCTCCCTAATATACCTC 444

RESULT 7
 AX200584 444 bp DNA linear PAT 29-AUG-2001
 LOCUS AX200584
 DEFINITION Sequence 214 from Patent WO0151633.
 ACCESSION AX200584
 VERSION AX200584.1 GI:15390404
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 444)
 AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
 Reed,S.G., Kalos,M.D., Fanger,G.R., Day,C.H., Retter,M.W.,
 Stolk,J.A., Skeiky,Y.A., Wang,A. and Mesgher,M.J.
 TITLE Compositions and methods for the therapy and diagnosis of prostate
 cancer
 JOURNAL Patent: WO 0151633-A 214 19-JUL-2001;
 CORIXA CORPORATION (US)
 FEATURES
 source Location/Qualifiers
 1..444
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 120 a 100 c 78 g 143 t 3 others
 ORIGIN

Query Match 22.5%; Score 425.8; DB 6; Length 444;
 Best Local Similarity 98.6%; Pred. No. 1.1e-97;
 Matches 438; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 688 accagaatccaatgctgaataattggtctcattatcccaagattcttgatgtcaag 747
 DB 1 ACCCAAAATCCAAATGCTGAATATTGGCTTCATTATCCAGATCTTGATGTCAAG 60

QY 748 gatttaattgtctcagcttgaggcactcaagtgagacctaaagatgcagccgag 807
 DB 61 GATTTAATGTGTGCTCAGCTTGCGCACTTCAGTTAGACCTAAGATGCCAGCCGAG 120

QY 808 ttataatgcagcaacaatattcagcgacacaggttattgacttgcccgagc 867
 DB 121 TTTATATATGCAACAATATTCAAGCGGACACAGATTATTTGAATGCGCCAGT 180

QY 868 tgaattcatcccatctgacttgagatccctatcatcagccagagattgaattac 927
 DB 181 TGAATTTCAATTCCTCATTTGAGATCCTTATCATCAGCCANAGATTTGAATTTTAC 240

QY 928 cccatgcactcttactctctcagagagcgagtggtgactataagcttgccaca- 966
 DB 241 CCTACGACTCTTACTCTCTGAGAGGCGCAAGTGTGTAAGCTTAAGCTTGCCACAT 300

QY 987 ttttttcttattcctcttgtagagatgtagatccatctgtagaaccacag 1046
 DB 301 TTTTTCCTTTATTCCTTTGTCAGAGATGCAATTCATTCCTANAAACCAACAG 360

QY 1047 agtgaatttacaatctcctataagagattgtaataaacttactactagttgccatt 1106
 DB 361 AGTGACTTTTACAAATTCCTATAGANATTGTGAATTAACCTTACTTATGTCGCAAT 420

QY 1107 acttgctctccctagatactc 1130
 DB 421 ACTTGTCTCTCCCTAATATACCTC 444

RESULT 8
 AX267240 444 bp DNA linear PAT 26-OCT-2001
 LOCUS AX267240
 DEFINITION Sequence 214 from Patent WO0173032.
 ACCESSION AX267240
 VERSION AX267240.1 GI:16516018
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (sites)
 AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
 Kalos,M.D., Fanger,G.R., Retter,M.W., Stolk,J.A., Day,C.H.,
 Vedrick,T.S., Carter,D., Li,S.X., Wang,A., Skeiky,Y.A., Hepler,W.T.
 and Henderson,R.A.
 TITLE Compositions and methods for the therapy and diagnosis of prostate
 cancer
 JOURNAL Patent: WO 0173032-A 214 04-OCT-2001;
 CORIXA CORPORATION (US)
 FEATURES
 source Location/Qualifiers
 1..444
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 120 a 100 c 78 g 143 t 3 others
 ORIGIN

QY 748 gatttaattgtctcagcttgaggcactcaagtgagacctaaagatgcagccgag 807
 DB 61 GATTTAATGTGTGCTCAGCTTGCGCACTTCAGTTAGACCTAAGATGCCAGCCGAG 120

QY 808 ttataatgcagcaacaatattcagcgacacaggttattgacttgcccgagc 867
 DB 121 TTTATATATGCAACAATATTCAAGCGGACACAGATTATTTGAATGCGCCAGT 180

QY 868 tgaattcatcccatctgacttgagatccctatcatcagccagagattgaattac 927
 DB 181 TGAATTTCAATTCCTCATTTGAGATCCTTATCATCAGCCANAGATTTGAATTTTAC 240

QY 928 cccatgcactcttactctctcagagagcgagtggtgactataagcttgccaca- 966
 DB 241 CCTACGACTCTTACTCTCTGAGAGGCGCAAGTGTGTAAGCTTAAGCTTGCCACAT 300

QY 987 ttttttcttattcctcttgtagagatgtagatccatctgtagaaccacag 1046
 DB 301 TTTTTCCTTTATTCCTTTGTCAGAGATGCAATTCATTCCTANAAACCAACAG 360

QY 1047 agtgaatttacaatctcctataagagattgtaataaacttactactagttgccatt 1106
 DB 361 AGTGACTTTTACAAATTCCTATAGANATTGTGAATTAACCTTACTTATGTCGCAAT 420

QY 1107 acttgctctccctagatactc 1130
 DB 421 ACTTGTCTCTCCCTAATATACCTC 444

RESULT 8
 AX267240 444 bp DNA linear PAT 26-OCT-2001
 LOCUS AX267240
 DEFINITION Sequence 214 from Patent WO0173032.
 ACCESSION AX267240
 VERSION AX267240.1 GI:16516018
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (sites)
 AUTHORS Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
 Kalos,M.D., Fanger,G.R., Retter,M.W., Stolk,J.A., Day,C.H.,
 Vedrick,T.S., Carter,D., Li,S.X., Wang,A., Skeiky,Y.A., Hepler,W.T.
 and Henderson,R.A.
 TITLE Compositions and methods for the therapy and diagnosis of prostate
 cancer
 JOURNAL Patent: WO 0173032-A 214 04-OCT-2001;
 CORIXA CORPORATION (US)
 FEATURES
 source Location/Qualifiers
 1..444
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 120 a 100 c 78 g 143 t 3 others
 ORIGIN

Query Match 22.5%; Score 425.8; DB 6; Length 444;
 Best Local Similarity 98.6%; Pred. No. 1.1e-97;
 Matches 438; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 688 accagaatccaatgctgaataattggtctcattatcccaagattcttgatgtcaag 747
 DB 1 ACCCAAAATCCAAATGCTGAATATTGGCTTCATTATCCAGATCTTGATGTCAAG 60

QY 748 gatttaattgtctcagcttgaggcactcaagtgagacctaaagatgcagccgag 807
 DB 61 GATTTAATGTGTGCTCAGCTTGCGCACTTCAGTTAGACCTAAGATGCCAGCCGAG 120

QY 808 ttataatgcagcaacaatattcagcgacacaggttattgacttgcccgagc 867
 DB 121 TTTATATATGCAACAATATTCAAGCGGACACAGATTATTTGAATGCGCCAGT 180

QY 868 tgaattcatcccatctgacttgagatccctatcatcagccagagattgaattac 927
 DB 181 TGAATTTCAATTCCTCATTTGAGATCCTTATCATCAGCCANAGATTTGAATTTTAC 240

QY 928 cccatgcactcttactctctcagagagcgagtggtgactataagcttgccaca- 966
 DB 241 CCTACGACTCTTACTCTCTGAGAGGCGCAAGTGTGTAAGCTTAAGCTTGCCACAT 300

QY 987 ttttttcttattcctcttgtagagatgtagatccatctgtagaaccacag 1046
 DB 301 TTTTTCCTTTATTCCTTTGTCAGAGATGCAATTCATTCCTANAAACCAACAG 360

QY 1047 agtgaatttacaatctcctataagagattgtaataaacttactactagttgccatt 1106
 DB 361 AGTGACTTTTACAAATTCCTATAGANATTGTGAATTAACCTTACTTATGTCGCAAT 420

QY 1107 acttgctctccctagatactc 1130
 DB 421 ACTTGTCTCTCCCTAATATACCTC 444

Oy	868	tgaattctatcccatgctgtggagtccttatactacgcacagagaagtgaatttac	927
Dd	181	TGAATTCTATCCCATGAGTGGAGTCCTTTATCACCACCCANAGAGATTGAAATTTTAC	240
Oy	928	cctaactcctttactctctgagagagcgacagtggtgtagctataagcttggccaca-	986
Dd	241	CCCTAGACTCTTTACTCTCTGTGAGAGGGCCAGTGCGTAGCTATTAAGCTTGGCCACAT	300
Oy	987	ctttcttcttatctctcttctgacagagatgatcatcacatagctagaaccaaacg	1046
Dd	301	TTTTTTTCTCTTATCTCTTGTGCAGAGATGCAGATTCATCATATGCTANMAAACCAACAG	360
Oy	1047	actgactttacaaaatttctctataagatgtgcaataaaaaccttaactaatgttccatt	1106
Dd	361	AGTGACTTTACAAAATTTCTATAGANNATYTGAAATAAACCCTTACCTTAAGTTGCCATT	420
Oy	1107	acttgtctccctagtlatactc 1130	
Dd	421	ACTTGTCTCTCCCTATATATACCTC 444	
RESULT	9		
LOCUS	AY029586	1971 bp	mRNA linear ROD 09-MAY-2001
DEFINITION	Mus musculus dudulin 2 mRNA, complete cds.		
ACCESSION	AY029586		
VERSION	AY029586.1 GI:14017412		
KEYWORDS	house mouse,		
SOURCE	Mus musculus		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 1971)		
TITLE	Serru,V., Lambdin,D., Lenoir,C., Manivet,P., Vaubourdlle,M.,		
JOURNAL	Kellermann,O. and Loric,S.		
REFERENCE	Molecular cloning and expression of mouse dudulin 2		
AUTHORS	2 (bases 1 to 1971)		
JOURNAL	Loric,S. and Serru,V.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (10-Apr-2001) Blochemistry A Laboratory, Saint-Antoine		
JOURNAL	University Hospital, 184 rue du Faubourg Saint-Antoine, Paris		
FEATURES	75012, France		
source	Location/Qualifiers		
CDS	1..1971		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	46..1590		
	/codon_start=1		
	/product="dudulin 2"		
	/protein_id="AAK50539.1"		
	/db_xref="GI:14017413"		
	/translation="MSGEMDKPLISRLVDSGLAEYPKAPVIGLSDGFARSILAT		
	TRLVGSGVVVGSFHPKRTAGLFPSLAQVTFQEBEAVSSPVIVAFREHYSLCSL		
	ADOLAGKLIVDVSNPTKEKHLOHROSNAEYIASLPACTVYKANVISAMALOAOPDL		
	GNOVLICGNDSKAKORYEMARAMEAGTPIDMGSLASAREVEATDRLPGSKPTLLI		
	ALGAFVCERYTNPIRDVLYQPIRKDEKFYMPISLVNTTIPCAVIYLSELYPGVI		
	AAALQURGRITYQRPMIDLHWLDHRKOIGLSFFPALHVALYSFCULPQRSHTYDIIV		
	NIAVKQYLANKSRLWEEVEVRMEIYSLGVLAGMSLAVTSIPSJANSLNKRERS		
	FVOSQIFGVALLISTMHITLVGMTRAFEENHYKFLIPTFTILLPLCVIIILAKGLFL		
	LPCISRLRTRTKIRRMWKDGAVKFMPLPDTHRGRRQATCEALEVMACGGPELGSLRS		
	GCCRAR"		
BASE COUNT	410 a 572 c 548 g 441 t		
ORIGIN			
Query Match	22.0%; Score 415.2; DB 10; Length 1971;		
Best Local Similarity	57.8%; Pred. No. 8.le-95;		
Matches 760; Conservative 0; Mismatches 548; Indels 6; Gaps 1;			
Oy 407	cactgtaggctgattggaagtgaagatttggccaactcttgaccactgactattatg 466		

Db	132	CAAAATGGGCAATCCTGGGCAATGGGGAATTTTGGCCGTTCCCTGGCCACACAGGCTGTGGTGG	191
Qy	467	atgcygctacatgbygbtcalaggaagtagaataactcaagttgtctctgtaattttctc	526
Db	192	CTCTGGCTTCAGATGTGTGGTGGGGAGCCCTTAACCCCAACACAGCGGCTGTCTTCC	251
Qy	527	tcatgbytagatgtaactcatcatagaaagtgtctcacaaaaaataataatattgt	586
Db	252	CTCCTTAACTCAAGTGAATTCCTCCAGGGAGAGCGGTGACCTCTCCAGAGTATATTGT	311
Qy	587	tgtctaacagagaacaatlatactccctgtgtgagaccgtgagacatctgtgtgtgtaa	646
Db	312	GGCCGCTGTCCGGGAGAGCACTACTCTCTACGTGTACACTTGTCTGTGACCACTTGGCTGGGAA	371
Qy	647	aatcctgatgaigtgagcaatlaacalgtagataaa-----ccagttaccaagtaacca	700
Db	372	GATCCCTGCTGGATGTAAAGCAACCCCAAGGAGAGAGATCTTGTAGCAACCCGCACTTAA	431
Qy	701	tgtctgaatattggtcttcattatccagaattctttgatgttcaagaagttaaagtgt	760
Db	432	TGCTAGTACCTGGCCCTCACCTCTTCTCGGTGTACACTGTGGTGAAGGCGCTTCAACGTAT	491
Qy	761	ctcagctgtggtgaccttcaatgtagaacctaaaggaatgcacgctgagattatataatgag	820
Db	492	CTCTCATGGGCGCTTAACAGGCTGTGCCAAGGATGGGAACAGCAGTGTCTATCTGGGG	551
Qy	821	caacaataatlaaagcgcgacaacaagytalttgaactgtgcgcgacagtgtgaattcatcc	880
Db	552	AAATGACAGCAAGGCCAAACAAGTAAGATGAGATGGACGCGCATGGTGTTCACACC	611
Qy	881	catgtactgtgatactcattcatcagccagagagatgtgaaatttaccctcagactct	940
Db	612	CTGTGACATGTGGGATTCCTGTGGCTTAAGGCGGGAGGTAAAGCCATACCCCTGCGCTCT	671
Qy	941	tactctcggagagagccagtggtgtgtagataaagcttggccacatttttttccctta	1000
Db	672	TGCATCCGGGAAGGTGGCCCAACCTCTCGGCACTGGGGGCTCTTTGTGTCTTACACTA	731
Qy	1001	ttcctttgtcagagatgltgaltcatccatagtctagaaaaccaacagatgactttacaa	1060
Db	732	CACCTTCAATCCGGGAGCGTTTCACAGCATATTCGGAAGATGAGAACAAATCTTCAAA	791
Qy	1061	aattctctatagatgtgtgaaataaaccttaccataagttgcattactttgtctcct	1120
Db	792	GATGGCCCTTGTCTGTGGTCAACACACACATACCTGTGGGCTTATGTGCTGTCTGTCCCT	851
Qy	1121	agttaactgtcaggtcttcttcgcagctgttgaacttaactttaacggaacaagtatag	1180
Db	852	GGTGTACTGCCCGGTGTGGTGGAGTGGCGTTCAGCTGCGGAGGGGACCAACTTACCA	911
Qy	1181	gagattctcacctgtgtgtgaaacctgtgtacaggttagaanaacagcttgaattactaag	1240
Db	912	GCGCTTCCCAACACTGCGCTGGAACCACTGTGCGAGCATGCCAAGCAGATTCGGGCTGTAG	971
Qy	1241	tttttctgtcctatgltgcatagltgtgcctacagcctctgtcctacgatgtgaaggtcaga	1300
Db	972	CTTCTTCTTCCGATGTGTCACACGCTCTTACACTTCTGCTGCGGCTGCGCGCTCCCA	1031
Qy	1301	gagatatttgtttctcaaatgtgcctacacagagcttcaagagttcaatattgaaactctg	1360
Db	1032	CCGCTACACACCTGTGTCTAATGTGCTGTGAAGCGAGGTCTCGGGCAACAGAGCCGCTCTG	1091
Qy	1361	gaatgaganaagaatttggagaaattgaaatgatactcccttggcataatgagagcttgg	1420
Db	1092	GCTTGAAGGACAAATCTCTGGAGATGAGATATACCTGTCTCCGTGGTGTGCTGGCCCTGGG	1151
Qy	1421	cttaacttccctctgtgcagltcaactctlatcccttcaagtgaagcaatgtctttaaactgag	1480
Db	1152	CATGTTGTGCGTGTGGCTGTTCACACTGTGCCGCTGCATTGTCTAATTCCTCAACTGGAA	1211
Qy	1481	agaatcagtttattcagctctaacacttggatagtgcgtctgtcctcataagtaacttcca	1540
Db	1212	GGAGTGAAGCTTCTGTGCATGTCCCACTGTGGCTTTCGTGGGCGCTGATATCTACACAAATGCA	1271

RESULT	10
AC026813	
LOCUS	205085 bp DNA linear HTG-27-OCT-2001
DEFINITION	Mus musculus chromosome 5 clone RP23-119M19 strain C57BL6/J, WORKING DRAFT SEQUENCE, 10 unordered pieces.
ACCESSION	AC026813
VERSION	AC026813.2 GI:16506398
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	house mouse.
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 205085) Akhter, N., Ayele, K., Beckstrom-Sternberg, S. M., Benjamin, B., Blakeslee, R. W., Bouffard, G. G., Breen, K., Binkley, C., Brooks, S., Dietrich, N. L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Ho, S.-L., Idol, J. R., Karlins, E., Latic, P., Lee-Lin, S.-O., Legaspi, R., Maduro, O. L., Maduro, V. B., Mastello, C., Mestian, S. D., McCluskey, J. C., McDowell, J., Pearson, R., Prasad, A., Shevchenko, Y., Stancitop, S., Thomas, J. W., Thomas, P. J., Touchman, J. W., Tsurgion, C., Vogt, J. L., Walker, M. A., Wetherby, K. D., Wiggins, L., Zhang, L.-H. and Green, E. D. NISC Comparative Sequencing Initiative Unpublished 2 (bases 1 to 205085) Green, E. D. Direct Submission Submitted (24-MAR-2001) NIH Intramural Sequencing Center, 8717 Government Circle, Gaithersburg, MD 20877, USA On Oct 27, 2001 this sequence version replaced gi:7321458.
TITLE	Genome Center
JOURNAL	
AUTHORS	Center: NIH Intramural Sequencing Center
REFERENCE	Center code: NISC Web site: http://www.nisc.nih.gov Contact: nisc.mouse@nhgri.nih.gov
COMMENT	----- Project Information ----- Center project name: xh Center clone name: 119M19 ----- Summary Statistics ----- Sequencing vector: plasmid; n/A; 100% of reads Chemistry: Dye-terminator Big Dye; 100% of reads Assembly program: Phrap; version 0.990319 Consensus quality: 199516 bases at least Q40 Consensus quality: 200556 bases at least Q30 Consensus quality: 201000 bases at least Q20 Insert size: 197000; agarose-fp Insert size: 192000; pulse-field-gel Insert size: 204185; sum-of-contris Quality coverage: 9.36x in Q20 bases; agarose-fp Quality coverage: 9.60x in Q20 bases; pulse-field-gel Quality coverage: 9.03x in Q20 bases; sum-of-contris ----- * NOTE: This is a 'working draft' sequence. It currently * consists of 10 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will

```
* be preserved.
```

	*	1	4428:	contig of 4428 bp in length
	*	4429	4528:	gap of unknown length
	*	4529	9553:	contig of 5025 bp in length
	*	4529	9653:	gap of unknown length
	*	9554	22746:	contig of 13093 bp in length
	*	9654	22846:	gap of unknown length
	*	22747	22847:	contig of 11108 bp in length
	*	22847	33954:	contig of 11108 bp in length
	*	33955	34054:	gap of unknown length
	*	34055	47831:	contig of 13777 bp in length
	*	47832	47931:	gap of unknown length
	*	47932	62010:	contig of 14079 bp in length
	*	62011	62110:	gap of unknown length
	*	62111	81789:	contig of 19679 bp in length
	*	81790	81889:	gap of unknown length
	*	81890	109919:	contig of 28030 bp in length
	*	109920	110019:	gap of unknown length
	*	110020	153235:	contig of 43216 bp in length
	*	153236	153335:	gap of unknown length
	*	153336	205085:	contig of 51750 bp in length.

FEATURES
source

1..205085
 organism="Mus musculus"
 strain="C57Bl6/J"
 /db_xref="taxon:10090"
 chromosome="5"
 clone="RP23-119M19"
 /clone_lib="RPCI mouse BAC library.23"
 1..4428
 misc_feature
 /note="assembly_fragment"
 4529..9553
 /note="assembly_fragment"
 9654..22746
 /note="assembly_fragment"
 22847..33954
 /note="assembly_fragment"
 34055..47831
 /note="assembly_fragment
 clone_end:r7
 vector_side:right"
 47932..62010
 /note="assembly_fragment"
 62111..81789
 /note="assembly_fragment"
 81890..109919
 /note="assembly_fragment"
 110020..153235
 /note="assembly_fragment"
 153336..205085
 /note="assembly_fragment
 clone_end:SP6
 vector_side:left"

BASE COUNT 61131 a 40716 c 41296 g 61026 t 916 others

ORIGIN

Query Match 21.7%; Score 411; DB 2; Length 205085;
Best Local Similarity 87.2%; Pred. No. 3.5e-93;
Matches 463; Conservative 0; Mismatch 65; Indels 3; Gaps 1.

QY	340	gccctcaagacgacctagtgaactgttlaaccatacgcataatgatcatcctaattgataggaa	339
DB	186621	MGGTATTCTCTGTGCTCTTTAAGCGAGTCCTCAACCTGGAAATCCATCTCTATGATGGGAA	186680
QY	280	aggaattcttgatgatcttggaagtgtcccgatcacatgaatcatcctaattgataggaa	339
DB	186681	GCCCTTAAGAAGCCT--GGAGACGCTTTTGCCATAAGGCATTAATGATGAACAAGACGCA	186737
QY	400	ggaagtcactcgttaagtgtgatctggaaatgtgaattttgcacaatccctgaccattgac	459
DB	186738	GGCAGGTACCCTGGGGGATAGGAAGAGGGGATTTTTGCCAAGTCTCTGACCATTTGGC	186797
QY	460	ttaataagatcgcatacattgacatcataagaatcaaaaatcciaaatlgttgttcgaat	519

```
||||| 186798 TTTATATGTGGCTACAGCTGGTCTATAGAACACAAATCCCAAGTTGCTACAGAT 186857
QY 520 ttcttcctcatgtgtagatgcatcatcatgaagaatgctctcacaacaaacataataa 579
Db 186858 TTTTCTCTACAGTGTAGAGTACCCACCATGAGATGCTTTACCAAAAAAATATTA 186917
QY 580 tatttgtctatacacagaagaacataataacccctctgtgtggacctgaagatctgtctg 639
Db 186918 TATTCTGGCTATCCATAGAGACATTAACATCTCTGTGGACCTGAGACATCTGCTTG 186977
QY 640 tgggttaaaatccctgatgagtagaataacatgaagataaaccagttaccagaatcca 699
Db 186978 TGGGCAAAATCCCTATTGATGATGACACACATGAGATGAACAGTACCAAGATCCA 187037
QY 700 atctgaataatgtgctcatatattccacagatctcttgatgttcaagaagattaatgtg 759
Db 187038 ATGCAGAGTACCTGGCTCTATTATTCCTCCGACATCTTGATTTCAAGATTTAATGTA 187097
QY 760 tctcagcttgggaactcagttaggaactaagaatgcacgcggcaggttt 810
Db 187098 TCTCAGCTTGGGACCTTACAGTACCTGACCAAGATGCCGCCAGGTAT 187148

RESULT 11
AC092404 227144 bp DNA linear HTG 04-JUL-2001
LOCUS Mus musculus chromosome 3 clone RP23-261D18 strain C57BL6/J,
DEFINITION WORKING DRAFT SEQUENCE, 3 unordered pieces.
AC092404
VERSION AC092404.1 GI:14595778
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 227144)
Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
Bouffard,G.G., Breun,K., Brinkley,C., Brooks,S., Dietrich,N.L.,
Gratlie,S., Guan,X., Gupta,J., Ho,S.-L., Idol,J.R., Karlins,E.,
Lee-Lin,S.-O., Legaspi,R., Lim,M., Maduro,Q.L., Maduro,V.B.,
Mastaglio,C., Mastrian,S.D., McLoskey,J.C., McDowell,J.,
Pearson,R., Prasad,A., Shevchenko,I., Snyder,B., Stantlirop,S.,
Thomas,J.W., Thomas,P.J., Touchman,J.W., Tsugeon,C., Vogt,J.L.,
Walker,M.A., Wetherby,K.D., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 227144)
Green,E.D.
Direct Submission
Submitted (04-JUL-2001) NIH Intramural Sequencing Center, 8717
Grovermont Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@nih.gov
Project information
-----
Center project name: gy
Center clone name: 261D18
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 225678 bases at least Q40
Consensus quality: 225859 bases at least Q40
Consensus quality: 226087 bases at least Q20
Insert size: 206000; agarose-fp
Insert size: 225000; pulse-field-gel
Insert size: 226944; sum-of-contigs
Quality coverage: 14.02x in Q20 bases; agarose-fp
Quality coverage: 12.84x in Q20 bases; pulse-field-gel
Quality coverage: 12.73x in Q20 bases; sum-of-contigs
```

```
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 26708: contig of 26708 bp in length
* 26809: gap of unknown length
* 26809: 109655: contig of 82847 bp in length
* 109656 109755: gap of unknown length
* 109756 227144: contig of 117389 bp in length.
Location/Qualifiers
1. 227144
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="5"
/clone="RP23-261D18"
/clone_1bp="RPCI mouse BAC library 23"
1. 26708
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
26809. 109655
/note="assembly_fragment"
clone_end:T7
vector_side:left"
109756. 227144
/note="assembly_fragment"
BASE COUNT 68114 a 46378 c 46791 g 65638 t 223 others
ORIGIN

Query Match 21.7%; Score 411; DB 2; Length 227144;
Best Local Similarity 87.2%; Pred. No. 3.6e-93;
Matches 463; Conservative 0; Mismatches 65; Indels 3; Gaps 1;

QY 280 aggatattcttggtgactcgtgaagtgctgcgtcatgaagaatcattcattgtagaa 339
Db 128222 AGGCATTTCTGGGCTTTAGACGAGTCTTACCATGAAATCCATCTATGATGGGAA 128163
QY 340 gccctaaagccttagtgaagactgtttaccctaagtcataaatgataatgaatgcaa 399
Db 128162 GCCCTAAGAGCCT---GGAGACGTTTTTGGCTATGCGATTAATGATCAAAAGACCAA 128106
QY 400 ggaagtcactgtgagtgtatggaagtgtgagatttcccaatccttgaccattgcac 459
Db 128105 GGCAGGTCCACCGTGGGGGTATAGGAAGTGGGGATTTTGGCAAGTCTCTGACCATTCGGC 128046
QY 460 ttattagatcggtatcatcatggtgtagaagtagaataaccgaattgtcttgaat 519
Db 128045 TTTATTAGATGTGGCTACAGCTGTGCTATGAGACCAAAATCCAACTTTGATCAGCAAT 127986
QY 520 ttcttcctcatgtgtagatgcatcatcatgaagaatgctctcacaacaaacataataa 579
Db 127985 TTTTCTCTACAGTGTAGAGTACCCACCATGAGATGCTTTACCAAAAAAATATTA 127926
QY 580 tatttgtctatacacagaagaacataataacccctctgtgtggacctgaagatctgtctg 639
Db 127925 TATTCTGGCTATCCATAGAGACATTAACATCTCTGTGGACCTGAGACATCTGCTTG 127866
QY 640 tgggttaaaatccctgatgagtagaataacatgaagataaaccagttaccagaatcca 699
Db 127865 TGGGCAAAATCCCTATTGATGATGACACACATGAGATGAACAGTACCAAGATCCA 127806
QY 700 atctgaataatgtgctcatatattccacagatctcttgatgttcaagaagattaatgtg 759
Db 127805 ATGCAGAGTACCTGGCTCTATTATTCCTCCGACATCTTGATTTCAAGATTTAATGTA 127746
QY 760 tctcagcttgggaactcagttaggaactaagaatgcacgcggcaggttt 810
```

Db 127745 TCTCAGCTTGGGCACTCAGTAGTCCCAAGATGCCAGCGCCAGGAT 127695

RESULT 12

AF335281 2707 bp mRNA Linear ROD 01-PEB-2001
DEFINITION Rattus norvegicus tumor suppressor pHyde mRNA, complete cds.
ACCESSION AF335281
VERSION AF335281.1 GI:12642828
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

1 (bases 1 to 2707)
Steiner, M.S., Zhang, X., Wang, Y. and Lu, Y.
Growth inhibition of prostate cancer by an adenovirus expressing a
novel tumor suppressor gene, pHyde

- JOURNAL

MEDLINE 20424188
Cancer Res. 60 (16), 4419-4425 (2000)
10969787

REFERENCE

2 (bases 1 to 2707)
Lu, Y., Rinaldy, A.R. and Steiner, M.S.
Direct Submission

Submitted (08-JAN-2001) Urology, University of Tennessee Health
Science Center, 956 Court Avenue, H220, Memphis, TN 38163, USA

FEATURES

Source

Location/Qualifiers

1..2707
/organism="Rattus norvegicus"
/strain="Dunning"
/db_xref="taxon:101116"
/cell_line="At-1"
/tissue_type="prostate cancer tumor"
26..1492
/codon_start=1
/product="tumor suppressor pHyde"
/protein_id="AAK00361.1"
/db_xref="GI:12642828"

CDS

BASE COUNT 574 a 788 c 715 g 630 t
ORIGIN

Query Match 21.5%; Score 406.6; DB 10; Length 2707;
Best Local Similarity 57.4%; Pred. No. 1.4e-92;
Matches 755; Conservative 0; Mismatches 554; Indels 6; Gaps 1;

Qy 407 cactgtagtctgattggaagtgtggaatttgcacaatcccttgacattgcaattatag 466
Db 112 CAAAGTGGGATCTCTGGGACGCGGATTTGCGCGTCCCGCCACACGCGCTGTGG 171
Qy 467 atgcgcatcatgtggtcatagaaagttagaattcctaagttgttctgaaattttcc 526
Db 172 CTCTGGCTTCTTTGGTGGTGGGGAAGCCGTAAACCCAAAGCAGCTGCGGCTTTCC 231
Qy 527 tcatgtgtagatgtcactcatcagaaagtgtctcacaataaataataatattgt 586
Db 232 CTCCTTACCCCAAGTGAATTCAGAGAGAGCGCTGAGCTCTCAGAGGTCATCTTGT 291
Qy 567 tgcatacacagagaattactcctctgtggagacctgtggaacctgtgtgtgttaa 646
Db 292 GGCCTGTTCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 351

Qy 647 aatccctgattgtagcaataacatagagataa-----ccagtaccagaatccaa 700
Db 352 GATCCTAGTGGATGTAGCAACCCACGAGAGAGAGCGCTTTCACACCGCCAGTCGAA 411
Qy 701 tgcataatattgcttcatatcccaagattccttgattgtlcaaaagattatgt 760
Db 412 CCGCGAGTACCTGGCTCCCTCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 471
Qy 761 ctcaagcttggcaactcagattaggaactaagatgccaagcggaagttatataag 820
Db 472 CTCTGATGGGCTTACAGGCTGGCCCAAGGAGTGGGAACAGCAGCTGCTCATTCGGG 531
Qy 821 caacaattcaagcgcgcaacaaggttattgaacttgcgcgcgaattgaatttcc 880
Db 532 TGACACGCTGGAAAGCAACACACACCTCTCAGAGATGGCGCGGCATGGTTACCCC 591
Qy 881 catgttgtagatcctcatcacaagcagaagattgaatttaaacctcagactct 940
Db 592 ACTGGACATGGGATCCCTGGCTGACGAGGAGGTAGAGGCCATCCCTGCGCTCT 651
Qy 941 tactctctgagagggcagtggtgtgtagctataaagcttggccacatttttcc 1000
Db 652 TCCATCTTGAAGGTCCACACCTCTGCGCTGGGCTTAAAGCACAAAGCTATGCTA 711
Qy 1001 ttccttgcagagatgtatcatcatatgctagaacaacaagagtagtacttaca 1060
Db 712 CAACCTTACATCCGGAGAGCTTCTACAGCGCTTACATCCGGAAGATGAGACAGTTCTA 771
Qy 1061 aattccatagagattgtgaataaacttaccatagttgtccattacttgcctc 1120
Db 772 GATGCCCTGTCTGTGTGTAACACACAGATACCTGTGTGCTTACGTGCTGTGCC 831
Qy 1121 agtatacccgcaagctctctgtgcaagcttcaacttataagcaaccaagatag 1180
Db 832 GGTATACCTGCTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 891
Qy 1181 gagattccacctgtgtgaaacctggttacaagttagaagaacagcttgattac 1240
Db 892 GCGCTTCCAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 951
Qy 1241 ttttcttgcctatgttgcattgttgcctacagcctctgtcctacagatgagaag 1300
Db 952 CTTTCTTTCGCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1011
Qy 1301 gggatattgttctcaaatgttatacagaaggttcaatgtcaaatatgaaacct 1360
Db 1012 CCGTATGATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1071
Qy 1361 gaatgagagaagatttggagaattgaatgtatatacctctgtgcaataatgag 1420
Db 1072 GGTGAGGAAGAAGTGTGGCGGATGAGATATACCTGCTGCTGCTGCTGCTGCTGCT 1131
Qy 1421 cttacttccctctgagcaagtcattccctcaagtgaagcaatgtcttcaactgg 1480
Db 1132 CATGCTGCTACGCTGCGGCGGTATCTGATCTCTTCAATGCAAACTCACTCACTG 1191
Qy 1481 agaattcaatttattcaatcactcaacttgatagatgctgcctgtctataagta 1540
Db 1192 GAGTGTACCTTTGTGCAATCTGCAACGCTGGGCTTGGCGCTGATCTGAGCAAT 1251
Qy 1541 tgttaattttagtgaagaaagccttcttgaagaagtaactaagatttataac 1600
Db 1252 CACCTCAGCTACAGCTGCTGAGACCGCTCTTTTGAAGAAACACTACAAGTTCTAC 1311
Qy 1601 accaacttctgtctgtctgttctgttcttccctcaatgttaattcttggtaaga 1660
Db 1312 ACCACATATCAGCTACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1371
Qy 1661 attccttcaatgataagcgaagaaactaagaatataaagagcttgggaag 1715
Db 1372 CCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1426

LOCUS	AX211372	3844 bp	DNA	linear	PAT 06-SEP-2001
DEFINITION	Sequence 18 from Patent W00159127.				
ACCESSION	AX211372				
VERSION	AX211372.1	GI:15523741			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 3844)				
AUTHORS	Tang,Y.T., Yue,H., Baughn,M.R., Yao,M.G., Bandman,O., Azimzai,Y., Tal,P., Gandhi,A.R., Ring,H.Z., Shih,L.L., Yang,J. and Pollicky,J.L. Drug metabolizing enzymes				
TITLE	Patent: WO 0159127-A 18.16-AUG-2001;				
JOURNAL	Inte Genomics, Inc. (US)				
FEATURES	Location/Qualifiers				
SOURCE	1..3844				
BASE COUNT	/organism="Homo sapiens" /db_xref="taxon:9606" /note="Incyte ID No: 5629033CB1"				
ORIGIN	862 a 1118 c 1049 g 815 t				
Query Match	20.1%; Score 381; DB 6; Length 3844;				
Best Local Similarity	56.2%; Pred. No. 5.1e-86;				
Matches	739; Conservative 0; Mismatches 570; Indels 6; Gaps 1;				
OY	407 cactgtaggtgtgatttgtagaagtgtgagatttgcacaatccttgaccattcgactatattag 466				
Db	138 CAAAGTGGGCAATCCCTGGAGGAGCGGAGCTTGGCCCGCTCCCTGGCCACAGCGCTGTGGG 197				
OY	467 atgcgcatactatgtgtcatagtagagaagaaatcctaagtttgcttctgaatttttc 526				
Db	198 CTCCTGGCTTCAAAAGTGGTGTGGGAGCCCGCAACCCCAAGCCACAGCAGGCTTTTC 257				
OY	527 tcatgttgtagatgtcactcatcatgaagatgtctctcaaaaaaataataatttgt 586				
Db	258 CTCACCGGCCCCAAGGACACTTTCACAGAGAGAGCAGTGAAGCTCCCGGAGGTGATCTTGT 317				
OY	587 tgcctacacagagaacattatactcctctgtgggaacttggagacttgcctgtgtgtgtaa 646				
Db	318 GCTCTGTTCCTGGAGCACTACTCTTCCACTGTGCACTGACAGCTGACAGCTGGCGGCA 377				
OY	647 aatcctgattgtagtgaagcaataaataagataaa-----ccagtaaccagaatccaa 700				
Db	378 GATCCTGGTGATGTGAGGAAACCTTACAGAGAGAAAGCAACCTTACAGATTCGTGATCCAA 437				
OY	701 tgcctgaataatttgcttaattatcccaagattctttagtttcaaaagattatgtgt 760				
Db	438 TGCTAGTAACTCGGCTCCTCTTCCCACTTGCACTGACAGTGGTCAAGGCGCTTCAAGTCA 497				
OY	761 ctcaagcttgggaactttagtagaactcctaagagatccagccggcgagttatataatag 820				
Db	498 CTCCTCCCTGGAGACCTTCGAGAGCTGGCCCAAGAGATGTAAACAGCAGGTGCCATGTGGG 557				
OY	821 caacaattcaagcgcagacaacagattatgaacttgcctgcgcagttgaatttattcc 880				
Db	558 TGACCAAGCCAGAGCAAGCGTGTCTGCGAGATGCGCTCGCCATAGGGCTTATGCC 617				
OY	881 catgaacttggatctctatatactagcagagagatttgaataattacccttagagctctt 940				
Db	618 CGTGGACATGGGATCCCTGGCTCAAGCTTGGGAGGTGAGGCGCAATGCCCTTGGCTCTCT 677				
OY	941 tactctcggagaaggcgacagtgtgttagtctataagctcttggccaacttttttcccta 1000				
Db	678 CCGCGGCTGGAGAGTGGCCACCGCTCTGGCCCTGGGGCTTTCGTCTGCTTGTATAGCTTA 737				
OY	1001 ttctcttgcagagatgtgtatcatcatatgtctagaacaaccaagagttgactttaaa 1060				
Db	738 CAACCTTCCTCCGGGACGCTTTCGACCCCTATGTGACGAAAGCCGAGCAAGTTCCTCA 797				

Oy	1061	aattccctataagaagcttgatggaataaaaccttaccctaagtgtgcattacttcctcct	1120
Dd	798	GCTGCCCCGGTCCGGGTCAACACCACACTGCCTGGCGGCTCAGTGCTGCTTCCT	857
Oy	1121	agtaacctcgcaagtcctctctgcagcgtcgttatacaactttatlcaggcaccaagtatag	1180
Dd	858	CSTGTACTTTGCCCGGGCGTGTGGGGCTGGCCCTGGACGCTGGCGGGCACCAATACCA	917
Oy	1181	gagattccaccttggtttggaaaacctggtltacagtgtgtgaanaaacagcttgtattacag	1240
Dd	918	GGCGTTCCTCCCGCACTGGCTGAGACCACTGGCTACAGCACCGACAAGCATGTCGGCTCTCAG	977
Oy	1241	ttttcttgcttagtgcagtcagtgtgtccttaaagcctcgtcatccagatgatgaagaagttaga	1300
Dd	978	CTTCTTTCGGCGCGCCCTGCACAGCCCTCTACAGCTTCTGTGTCGGCTGCGCGGGCCA	1037
Oy	1301	gagatatattgtttccacaatalgctgtcttacacgaaggtltaatgcaaatatgaaaaactcttg	1360
Dd	1038	CCGCTACGACACCTGGTCAACCTGGCAGTCAAGAGGTTCTGGCCAAACAGAGCCACTCTG	1097
Oy	1361	gaatgaggaagaagtttggagaattgaaatgtatatactcctttggcataatgagccttgg	1420
Dd	1098	GGTGAGGAGAGGAGGCTCTGGCGGATGGAATCTACCTCCCTGGGAGAGTCTGGCCCTGG	1157
Oy	1421	cttaacttcocctccctggcagatcaactctatcccttcagtgagcaatgctttaaacttggag	1480
Dd	1158	CACGTTGTCCCTGCTGCGCGCTGACCTCACTGCCCTTCATTGCAAACTGCTCACTGGAG	1217
Oy	1481	agaattcagtttattacagtcacaccttgatgataitgtcgtcgtcatabaagtacttcca	1540
Dd	1218	GGAGTTTCAGCTTTCGTTCAGTCTCTCACTGGGCTTTGTGGCCCTCGTGGAGACACTGCA	1277
Oy	1541	tgtttaattatgatgatgaaagcagacgtttggaggaagaagctacacagattttaacc	1600
Dd	1278	CACGCTCACCTCAAGCGCTGTGAGCCCGGCTTCGAGGAGAGCCGCTACAGTTCTACTTCGC	1337
Oy	1601	accaaacattgtctctgtcgtctgtttgtttccctcaatgtataattcgtggataatttt	1660
Dd	1338	TCCCACTTCACGCTCACGCTGCTGTGTGTCCTCGCTGCTCATCTCGGCCAAGCCCTGTT	1397
Oy	1661	attcctccatgataaagccgaagaactlaaaaaagaaatlaaaaaagcgttggaaaag	1715
Dd	1398	TCTCCTGCCCTGACATCAGCCGACAGACTCGCCAGATCCGGAGAGGCTGTGGAGAG	1452
RESULT_14			
LOCUS	AY029585	1845 bp	mRNA
DEFINITION	Homo sapiens dudulin 2 mRNA, complete cds.		
ACCESSION	AY029585		
VERSION	AY029585.1	GI:14017410	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 1845) Serru,V., Manivet,P., Lenoir,C., Eschwege,P., Lambin,D., Vaujourduille,M., Kellermann,O. and Loric,S. Dudulin 2, a new tumor antigen expressed in various human tumors Unpublished 2 (bases 1 to 1845) Loric,S. and Serru,V. Direct Submission Submitted (10-APR-2001) Biochemistry A Laboratory, Saint-Antoine University Hospital, 184 rue du Faubourg Saint-Antoine, Paris 75012, France		
AUTHORS			
JOURNAL			
TITLE			
FEATURES	Location/Qualifiers		
SOURCE	1..1845 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="2"		

/db_xref="taxon:9606"
 /clone="NT2RP4001138"
 /cell_line="NT2"
 /cell_type="teratocarcinoma"
 /clone_1ib="NT2RP4"
 /note="cloning vector: pME18SFL3; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction."
 135..1601
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAA91839.1"
 /db_xref="GI:7023107"
 /translation="MEEMDKPLISLHLVSDSSLAKEVDEAPKVGILSGDFARSLA
 TRLVSGFVVVGSRNPKRTARLPSAAQVTOEBAVSSPEVTFVAVFRRHYSLSCL
 SDLAGKILVDVSNPTDEHLOHRESNAEYLAISPTCTVYVAEPIVIAISMTLOAGPRD
 GNRQVPGIDQDEAPRAVEEMALAMFPMVDGSLASAEVEVEMPRLIPANKVPITL
 AGLFVCFYAVNPFVADVLOPYQESONKFEKFPVSVVNTTLCVAVYLSLYLREVL
 AAALDRGTATQRFPMWLDHMLQHRKQIGLSFCAALHLYSFCPLRKARIDLV
 NLAVROVLANKSHLWEEFVEMWEIYLSIGVIALGLSLAVTSLSPLNSINWREFS
 FVOSSIGFVALVSLTLHLTYGWTAFESRYKFPYLPFTFTLLVPCVVIILAKLFL
 LFCISRRRLARIRGWERESTIKFTLPTDHALAEKTSHV"

BASE COUNT 863 a 1148 c 1071 g 830 t
 ORIGIN

Query Match 20.1%; Score 379.4; DB 9; Length 3912;
 Best Local Similarity 56.1%; Pred. No. 1.3e-85;
 Matches 738; Conservative 0; Mismatches 571; Indels 6; Gaps 1;

QY 407 cactgtaggtgtgatttggaagtggaatttgcacaatccttgaccatttgacttatag 466
 DB 221 CAAAGTGGGCACTCCGTGGTGAAGCGGAGCTTTGCCCTCCCTGGCCACACAGCGCTGGGG 280
 QY 467 atcgagcatcatgtgtcatgaagaagtagaaatccctaaattgtctctgaatttttc 526
 DB 281 CTCTGGCTTCAAAAGTGTGGTGGGAGCGCCACACCCCAAGCAGACCGCGCTGATCC 340
 QY 527 tcatgtgtgtagatgtcatcatcatgaagaatgctctcaaaaaaataataatttgt 586
 DB 341 CTCAGCGGCCCAAGTCACTTCCAAAGAGGAGGAGTCCCGGAGGTCACTTTGT 400
 QY 587 tgtataacagaagaacattatattccctcctgttgagaccttgagacatctgtttgtgtaa 646
 DB 401 GCGTGTCTCCGGAGCACTACTCTTCACTGTCAGTCTCAAGTCCAGTCCGGGCAA 460
 QY 647 aatccctgattgattgagcaataacatgaagataaa-----ccagtaaccagaatccaa 700
 DB 461 GATCCGTGGTGGATGTAGACACCCCTACAGAGCAAGACACCTTCAGCATGTAGTCCA 520
 QY 701 tgcgaataatttgcttcatatttcccaattcttgattgtcacaagaatttaatttgt 760
 DB 521 TGCTGAGTACCTGGCTCTCCCTTCCCACTTGCACAGTGTCAAGGCTTCAATGTCAT 580
 QY 761 cttagcttgagcattcagtttagaagtagccagccgagagtttatatatgag 820
 DB 581 CTCTGCTGGACCTCAGAGCTGGCCCAAGGAGTGAACAGGAGGTGCCATCTGCGG 640
 QY 821 caacaatttcaagcagcaacaggtttttaacttgcgcgcagcttgattatctcc 880
 DB 641 TGACACGCCAGAAAGCCAGCGTCTGTCTGAGATGGCGCTCGCATGGGCTTCAATGCC 700
 QY 881 cattgacttggatcttcatcatcagccagagagattgaataatttaccctacagactct 940
 DB 701 CGTGGACATGGATCCCTGGCGGTACAGCTGGAGGTGAGGCGCATGCCCTGCGCTCCT 760
 QY 941 tactctcttgagagagccagtggtgtgtagctataaagcttggccacaattttttccttta 1000
 DB 761 CCGGCGCTGGAAAGTCCCAACCTGCTGGCCCTGCGGCTTGTGCTGCTTATGCTA 820
 QY 1001 ttccttgcagagatgtatcatcatatgtagaacaacagagtgactttacaa 1060
 DB 821 CAACCTTCGTCGGGACGTTCTGACGCCCTATGTGACAGAAAGCCAGAAACAGTTTTC 880

QY 1061 aattcctatagagattgttgaataaacccttaccatagatttgcaatttcttccct 1120
 DB 881 GCTGCCCGGTGCTGGTCAACACACACATGCGGTGGCTGACCTGCTGCTGCTACT 940
 QY 1121 agatatactgcagagcttcttcttgagagctgttatacattatagagcccaagtag 1180
 DB 941 CGTGTACTTCCCGGCTGTGGCGGCTGCTGACCTGACCTGCGGCGGCAACAGTACCA 1000
 QY 1181 gagatttcaaccttggttgaagaacctgttaccagtgtagaanaacagcttgattactaa 1240
 DB 1001 GCGCTTCCCGGACTGCTGGAGACCATGCTGACACACCGCAAGAGATGCGGCTGCTCAG 1060
 QY 1241 ttttcttgcatagttgcataagcttgcataagcttgcataagcttgcataagcttgcata 1300
 DB 1061 CTCTTCTGCGCGCGCGCTGACAGCGCTCTACAGCTTCTGCTGCGCGCGCGCGCA 1120
 QY 1301 gaatatatttcttcaacttgcataagcagcttgcataagcagcttgcataagcagcttgcata 1360
 DB 1121 CCGCTACGACCTGCTCAACCTGCAAGTCAAGCAAGGCTTGGCCAAAGACCAAGCACTCTG 1180
 QY 1361 gaatgaagaagaagtttgaagaattgaatglatatcctcttggcataatgaagccttgg 1420
 DB 1181 GGTGGAGAGAGAGTCTGGCGGATGGAGATCTACCTTCCCTGGAGATGCTGGCGCTCGG 1240
 QY 1421 cttaattccctcctgcagcttcaacttccatcccttcaagtagaagaatgtttaaactggag 1480
 DB 1241 CAGCTTGTCCCTGCTGCGCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGAC 1300
 QY 1481 agaatcagtttattatagcttcaacttgcataagcttgcataagcttgcataagcttgcata 1540
 DB 1301 GGAATTCAGCTTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1360
 QY 1541 tgtttaaattatgataagcaagccttlttgaagaagtagtactacagatttatacacc 1600
 DB 1361 CAGCTACCTTACGAGTGGAGCCCGGCTTGAAGAGAGCGGTACAAAGTTTCACTAGTCC 1420
 QY 1601 accaaacttgccttgccttgccttgccttgccttgccttgccttgccttgccttgccttgcct 1660
 DB 1421 TCCACACCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1480
 QY 1661 attcctcctatgataagcgaagaagcgaagaagcgaagaagcgaagaagcgaagaagcgaaga 1720
 DB 1481 TCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1540

Search completed: May 9, 2002, 02:12:49
 Job time: 18694 sec


```

: GENERAL INFORMATION:
: APPLICANT: Daniel E. Afar
: APPLICANT: Rene S. Hubert
: APPLICANT: Kahan Leong
: APPLICANT: Arthur B. Raitano
: APPLICANT: Douglas C. Saffran
: APPLICANT: Steve Chappell Mitchell
: TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
: TITLE OF INVENTION: EXPRESSED IN HUMAN CANCERS AND USES THEREOF
: FILE REFERENCE: 129.16US02
: CURRENT APPLICATION NUMBER: US/09/323,873A
: CURRENT FILING DATE: 1999-06-01
: PRIOR APPLICATION NUMBER: 60/087,520
: PRIOR FILING DATE: 1998-06-01
: PRIOR APPLICATION NUMBER: 60/091,183
: PRIOR FILING DATE: 1998-06-30
: NUMBER OF SEQ ID NOS: 32
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 339
: TYPE: PRT
: ORGANISM: Homo Sapiens
: US-09-323-873A-2

```

```

? APPLICATION NUMBER: US/09/083,521
? FILING DATE: Herewith
? CLASSIFICATION:
? ATTORNEY/AGENT INFORMATION:
? NAME: CERRONE, MICHAEL C.
? REGISTRATION NUMBER: 39,132
? REFERENCE/DOCKET NUMBER: PP-0527 US
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (650) 855-0555
? TELEFAX: (650) 845-4166
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 141 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? IMMEDIATE SOURCE:
? LIBRARY: PROSTUT10
? CLONE: 1691243
? US-09-083-521-1

```

```

OY      124 QYPEMAYSLAFDPSLIVKGFNVSA---WALOLGPKDASROYICSNNTQARQOYTE 180
        || | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db      116 VEEGSAAEQOLDILVPGATVVAFAFHLSAVNMMEHE-GP--LPEDVLVCGDSDRSADDEVAR 172

OY      181 LARQLNFIIP-IDGLSSAREIENLPRLFTLMR 213
        || : | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db      173 LAVAITGRPGIDGALRYAQRQLEPLTAVLINMR 206

RESULT          5
US-09-651-941-17
; Sequence 17, Application US/09651941
; Patient No. 6355470
; GENERAL INFORMATION:
; APPLICANT: ROUVIER, PIERRE E
; APPLICANT: WALTERS, DANA M
; APPLICANT: RAINIER, RUSS
; TITLE OF INVENTION: Genes Encoding Picric Acid Degradation
; FILE REFERENCE: BC1022 US NA
; CURRENT APPLICATION NUMBER: US/09/651,941
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,545
; PRIOR FILING DATE: 1999-10-03
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 17
; LENGTH: 227
; TYPE: PRF
; ORGANISM: Rhodococcus erythropolis HL PM-1
; US-09-651-941-17
```

```

Query Match          5.7%: Score 144.5; DB 4; Length 227;
Best Local Similarity 28.0%: Pred. No. 5.8e-07;
Matches 60; Conservative 33; Mismatches 88; Indels 33; Gaps 9;

OY      25 IKDARKVTVGVIGSGDFFKSLTIRLRCGYHNVIGISRPPKFASEFFPH-----VVD 75
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1 MKSSKIANVVG--GCGPQCKGLAYRFAAAGWPVYIGSRKAEKAEALAEVRRAGCGAYVS 58

OY      76 VTHEDALTKTNIIFFVAL-----HREHYTSLMDRLHLVGLKILI-----DVSNNMRIN 123
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      59 AADNAAADADCPILLLVVPYDGHRELVS---ELADIFAGKLTVSCVNPGLGFDKSGAYGID 115

OY      124 QYPSNMEYLASLPDPSLIYGFENVYSA---WALDQKDKSDROYIGTSNNIQAQOYIE 180
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      116 VEESSAEQLDRLVPGAVVAAAFHLLSAVNLMEHE-GP--LPEDVLVCGDDRSARDEVAR 172

OY      181 LARQLNFIIP-IDGLSSAREIENLPLRLFTLMR 213
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      173 LAVAITGRPIDGGALRVARQLEPLTAVLVINVR 206

RESULT      6
US-08-724-974A-2
: Sequence 2, Application US/08724974A
: Patent No. 5912335
: GENERAL INFORMATION:
: APPLICANT: Derk J. Bergsma, Catherine E. Ellis
: TITLE OF INVENTION: A No. 5912335el G-Protein Coupled Receptor
: TITLE OF INVENTION: HUVCT36
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SmithKline Beecham Corporation
: STREET: 709 Swedeland Road, P.O. Box 1539
: CITY: King of Prussia
: STATE: PA
: COUNTRY: USA
: ZIP: 19406-0939
: COMPUTER READABLE FORM:
: MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
: COMPUTER: IBM 486

```

```

? OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
? SOFTWARE: WORDPERFECT 5.1
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/724,974A
? FILING DATE: October 3, 1996
? CLASSIFICATION: 424
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: William T. Han
? REGISTRATION NUMBER: 34,344
? REFERENCE/DOCKET NUMBER: ARG50022
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 610 270 5024
? TELEFAX: 610 270 5090
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 531
? TYPE: Amino Acid
? TOPOLOGY: Linear
? US-08-724-974A-2

```

[illegible]

```

: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: William T. Han
: REGISTRATION NUMBER: 34,344
: REFERENCE/DOCKET NUMBER: AIG50022
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 610 270 5024
: TELEFAX: 610 270 5090
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 365
: TYPE: Amino Acid
: TOPOLOGY: Linear
: US-08-724-974A-3

Query Match          4.4%; Score 112.5; DB 2; Length 365;
Best Local Similarity 20.6%; Pred.No. 0.0025;
Matches 56; Conservative 46; Mismatches 93; Indels 77; Gaps 11

Qy      255 VNKLT-PIVAITLLSLVYLAGILAAAYOLUYGTGKYRRFPMTWTWLOCKOGLISFFPA 313
       19 INGTLPVVYYIVL-----VGFRPANKLSIYFG-----YLQ--TKANETAGVIYCNLT 64
Qy      314 MYVAVASICLRPRRSERYELFNMAVOOVHANIENTSNWEERWKRIEMYSFGIMSIGLSL 373
       65 VADLFYICSLRPW-----LQVLQHDNMSHGDLSGVCGILLYENIYL-SVGFLCC 114
Qy      374 LAVTSPISNMNLNMRREFSIQSTLGVA-----LLISTF----- 408
       115 ISVDRLAVAHPRRHQPFRTLKAAVRVTYIVAKELLTSIFYLMHEVIEDENQRVCPE 174
Qy      409 HVLITYGMKRAFEERYRFTPNPFVALVLPST-----VIIG 445
       175 HPIQAMORAT--NYTRFLVGFLPFCILLASTYOGLIRAVRSHGTOKSRKDIOIKRWLS 232
Qy      446 KIIFLPCL-SRKLRINKGWKSQFLREGI 475
       Db      233 TVVIFLACFLPYHVLLLVRSVMEASODFAKGV 264

RESULT 8
US-09-323-873A-20
: Sequence 20, Application US/09323873A
: Patent No. 6329503
: GENERAL INFORMATION:
: APPLICANT: Daniel E. Afar
: APPLICANT: Rene S. Hubert
: APPLICANT: Kahan Leong
: APPLICANT: Arthur B. Raitano
: APPLICANT: Douglas C. Saffran
: APPLICANT: Steve Chappell Mitchell
: TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
: FILE REFERENCE: 129.16USU2
: CURRENT APPLICATION NUMBER: US/09/323,873A
: PRIOR FILING DATE: 1998-06-01
: PRIOR APPLICATION NUMBER: 60/091,183
: PRIOR FILING DATE: 1998-06-30
: NUMBER OF SEQ ID NOS: 32
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 20
: LENGTH: 34
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: STEAP-1 PEPTIDE
: US-09-323-873A-20
```

```

Query Match Similarity      4.2%   Score 107; DB 4; Length 34;
Best Local similarity       58.8%; Pred. No. 0.00025;
Matches    20; Conservative     5; Mismatches     9; Indels         0; Gaps          0

Oy      326 RRSRYFLNLNMAVQOVHANIENSMNEEYWRTEM 359
        ||| |  ||||| | :|:| | :|||:|:|
Db       1 RRSRYKILLNMAVQQVOQNKDAMIEHDVWMEI 34

RESULT      9
US-08-487-886-2
Sequence 2, Application US/08487886
Patent No. 5744448
GENERAL INFORMATION:
APPLICANT: Kelton, Christie Ann
APPLICANT: Schweickhardt, Rene Lynn
APPLICANT: Cheng, Shirley Yui Yen
APPLICANT: Nugent, No. 5744448een Petrice
TITLE OF INVENTION: Human Follicle Stimulating
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Stephan P. Williams,
ADDRESSER: Ares-Serono, Inc.
STREET: Exchange Place, 37th floor
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" diskette, 1.44 MB, high density
COMPUTER: IBM PS/2, model 55 SX
OPERATING SYSTEM: MS-DOS version 4.0
SOFTWARE: VAX/VMS Massll via Kermit to IBM MS-DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,886
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/670,085
FILING DATE: 15-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Williams, Stephen P.
REGISTRATION NUMBER: 28546
REFERENCE/DOCKET NUMBER: US/252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 723-1300
TELEFAX: (617) 723-8923
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 695
TYPE: Amino acid
TOPOLOGY: Linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: signal sequence
LOCATION: -17 to -1
IDENTIFICATION METHOD: hydrophobic
FEATURE:
NAME/KEY: putative amino-terminal extracellular domain
LOCATION: 1 to 349
IDENTIFICATION METHOD: similarity with other
IDENTIFICATION METHOD: dimeric glycoprotein receptor extracellular
IDENTIFICATION METHOD: domains, hydrophilic
FEATURE:
NAME/KEY: transmembrane domain
LOCATION: 350 to 613
IDENTIFICATION METHOD: similarity to other G
IDENTIFICATION METHOD: protein-coupled receptor transmembrane domains
FEATURE:
NAME/KEY: putative transmembrane region I

```

LOCATION: 350 to 370 similarity to other G
IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
FEATURE:
NAME/KEY: putative transmembrane region II
LOCATION: 382 to 404 similarity to other G
IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
FEATURE:
NAME/KEY: putative transmembrane region III
LOCATION: 427 to 448 similarity to other G
IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
FEATURE:
NAME/KEY: putative transmembrane region IV
LOCATION: 469 to 491 similarity to other G
IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
FEATURE:
NAME/KEY: putative transmembrane region V
LOCATION: 512 to 533 similarity to other G
IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
FEATURE:
NAME/KEY: putative transmembrane region VI
LOCATION: 557 to 580 similarity to other G
IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
FEATURE:
NAME/KEY: putative transmembrane region VII
LOCATION: 592 to 613 similarity to other G
IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
FEATURE:
NAME/KEY: putative carboxy-terminal intracellular
NAME/KEY: domain
LOCATION: 614 to 678
US-08-487-886-2

Query Match 4.2%; Score 106; DB 1; Length 695;
Best Local Similarity 16.4%; Pred. No. 0.031;
Matches 84; Conservative 81; Mismatches 148; Indels 198; Gaps 18;

QY 79 HEDALTKNIIIVAIHREHYTSLMDRLKLVG-----KIILIDVSNMR 121
DB 98 HEIRIEKANNL-LYINPEFQMLPNLOYLLISNTGIKHLPDVHKIHSLOKVLIDIDNIN 156
QY 122 INQVESNAEYLASFPPDLIVKGFNVVSAMALQGPDKASROYICSNINQARQVIEL 181
DB 157 IH-----TIERNSEVLSESVILML-----NKGIOEINCA----- 189
QY 182 AOLNFIPIIDGLSSAREIENLPLRLFTLMRGPPVVAISLATFFELVSFVDVIHPYAR 241
DB 190 ---FNGTQDELINLSDNNLELPNDVFGASGPVILDISPTRISLPSYGLLENKRLA 246
QY 242 NQGSDFKIPRIIVKTL-PIVAITLLSLVY----- 271
DB 247 RSTYMLKLP-----TEKVLALMEASLTYPSCAFANMRQISELHPICNKSILRQ 300
QY 272 -----LAGLAAAYQOLYGYRRF-----PPULETWLOCROJL 305
DB 301 VQYMTQTRQGRSSLAEDNSSYSRGEDMYTEFDVDCNEVVDYICSPFPAFNCEDIM 360
QY 306 G-----LISFFAMVHA-----YSLCLPMRSERYLFLNMAVQOVHANEN 347

DB 361 GYNILRLWIFSLIATGNIIVLITTSQYKLTVP-----RFLMCNLAFAAD----- 408
QY 348 SWESEEWNIEMISGIMSLSGLSLAATISVSNAIMNR-----FSFIQSTL 398
DB 409 -----LCIGIYLLITASVDIHKSOYHNVAIDMORGACDAGFTVASEL 455
QY 399 G---YVALLISPEH-----VLIVGMRAREEYRYPTPNFV-LA 435
DB 456 SVYTLTAITLERHNTTHAMQDCKQVLRHAASVVMGWFPAALFPIFGISSTYKVS 515
QY 436 LVLP-----SVILGKILFLPC 453
DB 516 ICLPMDISPLSOLYMSILVLNVLAFVVIC 546
RESULT 10
US-08-482-855-2
Sequence 2, Application US/08482855
Patent No. 6121016
GENERAL INFORMATION:
APPLICANT: Kelton, Christie Ann
APPLICANT: Schweickhardt, Rene Lynn
APPLICANT: Cheng, Shirley Vui Yen
APPLICANT: Nugent, No. 6121016een Patrice
TITLE OF INVENTION: Human Follicle Stimulating
TITLE OF INVENTION: Hormone Receptor
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Stephan P. Williams,
ADDRESSEE: Ares-Serono, Inc.
STREET: Exchange Place, 37th floor
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" diskette, 1.44 MB, high density
COMPUTER: IBM PS/2, model 55 SX
OPERATING SYSTEM: MS-DOS version 4.0
SOFTWARE: VAX/VMS Mass11 via Kermit to IBM MS-DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,855
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/670,085
FILING DATE: 15-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Williams, Stephan P.
REGISTRATION NUMBER: 28546
REFERENCE/DOCKET NUMBER: US/252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 723-1300
TELEFAX: (617) 723-8923
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 695
TYPE: Amino acid
TOPOLOGY: Linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: signal sequence
LOCATION: -17 to -1
IDENTIFICATION METHOD: hydrophobic
FEATURE:
NAME/KEY: putative amino-terminal extracellular domain
LOCATION: 1 to 349
IDENTIFICATION METHOD: similarity with other
IDENTIFICATION METHOD: dimeric glycoprotein receptor extracellular
FEATURE:
NAME/KEY: transmembrane domain
LOCATION: 350 to 613

```

1 IDENTIFICATION METHOD: similarity to other G
2 IDENTIFICATION METHOD: protein-coupled receptor transmembrane domains
3 FEATURE:
4 NAME/KEY: putative transmembrane region I
5 LOCATION: 350 to 370
6 IDENTIFICATION METHOD: similarity to other G
7 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions
8 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
9 FEATURE:
10 NAME/KEY: putative transmembrane region II
11 LOCATION: 382 to 404
12 IDENTIFICATION METHOD: similarity to other G
13 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions
14 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
15 FEATURE:
16 NAME/KEY: putative transmembrane region III
17 LOCATION: 427 to 448
18 IDENTIFICATION METHOD: similarity to other G
19 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions
20 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
21 FEATURE:
22 NAME/KEY: putative transmembrane region IV
23 LOCATION: 469 to 491
24 IDENTIFICATION METHOD: similarity to other G
25 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions
26 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
27 FEATURE:
28 NAME/KEY: putative transmembrane region V
29 LOCATION: 512 to 533
30 IDENTIFICATION METHOD: similarity to other G
31 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions
32 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
33 FEATURE:
34 NAME/KEY: putative transmembrane region VI
35 LOCATION: 557 to 580
36 IDENTIFICATION METHOD: similarity to other G
37 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions
38 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
39 FEATURE:
40 NAME/KEY: putative transmembrane region VII
41 LOCATION: 592 to 613
42 IDENTIFICATION METHOD: similarity to other G
43 IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions
44 IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
45 FEATURE:
46 NAME/KEY: putative carboxy-terminal intracellular
47 NAME/KEY: domain
48 LOCATION: 614 to 678
49
50 US-08-482-855-2

```

[illegible]

```

Db      301  VDMYQTFGGGSSLAIEDNESSYSGFDMWTTEPFYDLCNEVVDTCSPKPDAAFNPCEDIM 360
Qy      306  G-----LISFFAAVHA-----YSLCEMRKSERFLFNMAVQOYVHANEN 347
Db      361  GYNILRFYIMFISILATGNIIYLVILTTSOYKLTVP-----RLMGNLAFAD----- 408
Qy      348  SMNEEYVRIMETYSFGIMSLGLSLTAVTSPESNALMRE-----ESFIOSTL 398
Db      409  -----LCIGIYLLINISVDIHKRSQYHNVAIDMQGAGCDAAGFTTFASEL 455
Qy      399  G---YVALLISTFH-----VLYGMRAREEERYRRTTPPNFY-LA 435
Db      456  SVYTLATTLTTERWHTITHAMODCKVOLRIHAASVYMWGMIFAAAPALPFIIGISYMKVS 515
Qy      436  LVLP-----SIYILGKIILFLPC 453
Db      516  ICLPPADIDSPISQLYVMSLVLANLVAFVVIC 546

```

```

RESULT 11
US-09-316-083-3
: Sequence 3, Application US/09316083A
: Patent No. 6280942
: GENERAL INFORMATION:
: APPLICANT: The Institute of Physical and Chemical Research
: TITLE OF INVENTION: Endonuclease
: FILE REFERENCE: PH-651
: CURRENT APPLICATION NUMBER: US/09/316,083A
: CURRENT FILING DATE: 1999-05-20
: EARLIER APPLICATION NUMBER: JP98/141861
: EARLIER FILING DATE: 1998-05-22
: NUMBER OF SEQ ID NOS: 38
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 3
: LENGTH: 476
: TYPE: PRT
: ORGANISM: Saccharomyces cerevisiae
: US-09-316-083-3

```

	Query Match	4.0%; Score 101; DB 4; Length 476;
	Best Local Similarity	20.7%; Pred. No. 0.057;
Matches	52; Conservative	48; Mismatches 85; Indels 66; Gaps 12.
OY	62 NPKFASEFPFHVVDTTHEDALFTKTNI-----IFVAIHREHYT-----SLMD	103
Dd	254 NPVEFNATSIIN-----KTNLAKKEIFTNINYLNYSDYKINOINNHIPIYNYLK	302
OY	104 LRHLVGLKLLIDVSNMNRINQYPESNAEVLSTF--PSLIYKGKNVVSAMMLQLGPKDA	161
Dd	303 INNKLPINIMDIKKNNYLACGTADGSFLSSWMPKDTLFRKM-----	347
OY	162 SROYVCISNNIOAROOVELARQLNFIEDLSLSAREENLPLRLFTLMRGPPVVAIS	221
Dd	348 -RPSYVIS-QYETRKELLYIQE---SPDL-SISNVKKVGARKLKDRLFRTDDELAK	400
OY	222 LATEPELYSFPRDVYHPARNOOSFYRIIP----LEIVNTKLPIYAITLISLVYLAGLA	277
Dd	401 -----FIYYF--DKPLPLDNKQFMYIKERENTFKSYMMNRNVGLVLTSE--YINNKI	451
OY	278 AAYOLUYGTRY	288
Dd	452 DNIYIIYNKY	462

RESULT 12
US-08-460-576-2
Sequence 2, Application US/08460576
Patent No. 6033903
GENERAL INFORMATION:
APPLICANT: SISK, William P.
TITLE OF INVENTION: A METHOD OF EXPRESSING AND SECRETING
TITLE OF INVENTION: SOLUBLE EXTRACELLULAR DOMAINS OF HUMAN GONADOTROPIN

Thu May 9 08:44:29 2002

us-09-802-520-1.raii

Page 7

TITLE OF INVENTION: HORMONE RECEPTORS
 NUMBER OF SEQUENCES: 18
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: BROWDY AND NEWMARK
 STREET: 419 Seventh Street, N.W., Suite 300
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20004
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/460,576
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: YUN, Allen C.
 REGISTRATION NUMBER: 37,971
 REFERENCE/DOCKET NUMBER: SISK-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-628-5197
 TELEFAX: 202-737-3528
 TELEX: 248633
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 390 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-460-576-2

[illegible]

```

1      ZIP: 02109
2
3      COMPUTER READABLE FORM:
4
5      MEDIUM TYPE: Floppy disk
6
7      COMPUTER: IBM PC compatible
8      OPERATING SYSTEM: PC-DOS/MS-DOS
9      SOFTWARE: PatentIn Release #1.0, Version #1.25
10
11     CURRENT APPLICATION DATA:
12
13     APPLICATION NUMBER: US/07/757,342D
14
15     FILING DATE: 10-Sep-1991
16
17     CLASSIFICATION: <Unknown>
18
19     ATTORNEY/AGENT INFORMATION:
20
21     NAME: BUCKLEY, Linda M.
22     REGISTRATION NUMBER: 31003
23     REFERENCE/DOCKET NUMBER: 41226
24
25     TELECOMMUNICATION INFORMATION:
26
27     TELEPHONE: (617)523-3400
28     TELEFAX: (617)523-6440
29     TELEX: 200291 STRE UR
30
31     INFORMATION FOR SEQ ID NO: 6:
32
33     SEQUENCE CHARACTERISTICS:
34
35     LENGTH: 692 amino acids
36     TYPE: amino acid
37     TOPOLOGY: linear
38
39     MOLECULE TYPE: protein
40
41     SEQUENCE DESCRIPTION: SEQ ID NO: 6:
42
43     GS-07-757-342D-6

```

[illegible]

```

GENERAL INFORMATION:
APPLICANT: Londenborough, John
APPLICANT: Vuorio, Outi
TITLE OF INVENTION: A method to increase the trehalose content
of organisms by transforming them with the
TITLE OF INVENTION: structural genes for the short and long chains
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: ALKO Ltd.
CITY: PO Box 350
STATE: Helsinki
COUNTRY: Finland
ZIP: SF-00101
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 720 Kb
COMPUTER: IBM PC/Xt/AT
OPERATING SYSTEM: PC-DOS
SOFTWARE: WP5.1 file exported as DOS text file
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/841,997A
FILING DATE: 19920228
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/836,021
FILING DATE: February 14, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Mary E. Gormley
REGISTRATION NUMBER: 34409
REFERENCE/DOCKET NUMBER: 920085A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)659-2930
TELEX: (202)887-0357
TELEX: 440142
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 495 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: Polypeptide
HYPOTHETICAL: yes
US-07-841-997A-2

Query Match          3.5%; Score 89.5; DB 1; Length 495;
Best Local Similarity 17.8%; Pred. No. 0.94;
Matches      83; Conservative    53; Mismatches   145; Indels   185; Gaps     20.

QY      80 EDALTKNTIIFVAIHRE---HYTS-----LMDLRHLVGTKILIDVSNMRINQYDESNA 130
       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      80 KDLEKFNAAVPFLSDEIDLHYNGFSNSILMPLFHHGGEINF D--ENAMFG-YNEANQ 136
       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      131 -----EYLASFPPDSLIYKF-----NVSASALDLGRKDSARQY 166
       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      137 TTTNEIAKTMMNHDLIWDYDHLMLVPEMLRVKIHKKOLQNKYGVGMFLR--TPPSSSELY 194
       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      167 ICSNNIQAROOVIE-----LARLNF----- 187
       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      195 ---RLLPKROELIKGLVSCDLVGFTTYDARRFLSSVQRVLAVNLPLNGVEYQGRVANG 251
       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      188 --IPIDIGSLSSARE-----IENLPRLRLTL---- 211
       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      252 APFIGIDVDKFTDGLKKESVQKRIOQLKETFGCKKIYGVDRLDYTIKGVPQKLHAMEVPL 311
       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      212 ----WRGPVV--VAISLATFFELYSTFYRDVIHPYARNQSDDFK--IPIEIYNKLTLP 260
       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      312 NEHPMRGKVYLVQYAVPSRGDVEEYOYLRSVVNELVGRINGOFGTVEEVPTPHFMKSIP 371
       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      261 IVATILLIVYLAGLLAAAYOLYGTGYKRYRRPPIWLETWLQCRKQGLISFPFAMVHAYS 320
       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      372 FEELISLAVSDVCIVSSI-----RDGMALVSY-----EYI 402

```

```

Yy      321  ICLIMRSEBRL--FLNMXXOOVHANENSNWNEEYWRIMETYSFGIMSLGLSLA VTS 378
Db      403  ACQOEKKSLLSEFTGGAOSLNGAIIIVNFRNDDLSL-----AINEALT 447
Qy      379  IPSVSNALNWRSEFSIOSTGLGVALLSTFHVLIYGNKRAFEERY 424
Db      448  LPDYKKEVNW-----EKLYYISKYTSAF-----WGENFVHELX 481

RESULT 15
US-08-290-301-2
Sequence 2, Application US/08290301
Patent No. 5792921
GENERAL INFORMATION:
APPLICANT: Londesborough, John
APPLICANT: Tunnela, Outi
APPLICANT: Palva, Tuplio
APPLICANT: Holmstrom, Kjell-Ove
APPLICANT: Wellin, Bjorn
APPLICANT: Mandel, Abul
TITLE OF INVENTION: Increasing the trehalose content
TITLE OF INVENTION: of organisms by transforming them with combinations of
TITLE OF INVENTION: the structural genes for trehalose synthase.
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: ALKO Ltd.
STREET: PO Box 350
CITY: Helsinki
STATE:
COUNTRY: Finland
ZIP: SF-00101
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 Inch, 720 KB
COMPUTER: IBM PC/XT/AT
OPERATING SYSTEM: PC-DOS
SOFTWARE: WP5.1 file exported as DOS text file
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,301
FILING DATE: 15 August 1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FI 943133
FILING DATE: 29 June 1994
APPLICATION NUMBER: PCT/FR93/00049
FILING DATE: 15 February 1993
APPLICATION NUMBER: 07/841,997
FILING DATE: 28 February 1992
APPLICATION NUMBER: 07/836,021
FILING DATE: 14 February 1992
ATTORNEY/AGENT INFORMATION:
NAME: Kubovcik, Ronald J.
NAME: Lydon, James C.
REGISTRATION NUMBER: 25,401
REGISTRATION NUMBER: 30,082
REFERENCE/DOCKET NUMBER: LAIN-001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 467-6300
TELEFAX: (202) 466-2006
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 495 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Polypeptide
HYDROTHERICAL: yes
US-08-290-301-2

Query Match 3.5%; Score 89.5; DB 1; Length 495;
Best Local Similarity 17.8%; Pred. No. 0.94;
Matches 83; Conservative 53; Mismatches 145; Indels 185; Gaps 20;
80 EDALTKTNIIFVAIHR-----HTS-----LMDIRHLVGVKILIDVSNMRRINQYRESNA 130

```



```
Db      80 KDLLEKFNVAVPITLSEIDELAHNGFSNSILMPLFHYHGEINF--ENAMG-YNENQ 136
QY      131 -----EYLAFLPDSLIYKF-----NVYSAWALQIGPKDASROY 166
Db      137 TETNEIAKTMNNDLIWHDYHMLVPEMLRVKIHKKOLQNVKGMFLH--TPPSSEIY 194
QY      167 ICSNNIOARQOYIE-----LAROLF----- 187
Db      195 ---RIIPVROEIIKGYLSCDVLGFEHTYDARHFLSSVQRYLVNNTLPNGVEYQGRFVNG 251
QY      188 ---IPIDGLSSARE-----IENLPLRLFTL---- 211
Db      252 AEPIDGIDVKFTDGLKKEVSQKRIQOLKFTFKCKIIVGVDRDLYIKGVPOKRLHAMEVEL 311
QY      212 ----WRGPVY--VAISLATFFELYSFVADVIHPIARNOQSDPK---IPLEIVNKTLP 260
Db      312 NEHPEMRGKVVLVQVAVPSPRGDVEEYOYLRSVNELVGRINGOFGTVEFVPIHFHKSIP 371
QY      261 IVAITLLSVYLAGLLAAAYQLYYGTKYRRFPPELFTWLQCRKOLGLSFEFFAMHVAYS 320
Db      372 FEELISLVAVSDVCLVST-----RDGMNLVSY-----EYI 402
QY      321 LCLPMRSEERYL--FLNMAYQOVHANIENSWNEEYVRIEMTISFGIMSLGILSLAYTS 378
Db      403 ACOEEKKGSLLISEFTGAQSLNGALITVNPWNTDLS-----AINEALT 447
QY      379 IPSVSNALMWRERSFIQSTLGYVALLISTFHVILITGMKRAFEERY 424
Db      448 LPDYKKEVNW-----EKLYKISKYTSAF-----WGENFVHELY 481
```

Search completed: May 9, 2002, 03:15:32
Job time: 11551 sec

THIS PAGE BLANK (USPTO)

PD 04-OCT-2001.
 XX 23-MAR-2001; 2001WO-US09410.
 XX 24-MAR-2000; 2000US-191929P.
 XX (SAAT/) SAATC1OGLU F.
 PA Saatioglu F.
 PI
 XX WPI: 2001-662926/76.
 DR N-PSDB; AAS15793, AAS15801, AAS15802.
 XX
 PT New polynucleotide for the diagnosis, prevention and treatment for
 PT prostate and testis disorders, particularly prostate cancer, comprises
 PT prostate-specific or testis-specific nucleic acids
 XX
 PS Claim 1; Fig 4B; 114pp; English.
 XX
 CC The invention relates to substantially pure prostate-specific or
 CC testis-specific polypeptides and the nucleic acids encoding them.
 CC Also included are vectors and host cells expressing the proteins, a
 CC transgenic animal expressing the protein, antibodies against the
 CC proteins, probes for detecting the nucleic acids, antisense molecules
 CC for the nucleic acids and methods of isolating modulators of the
 CC proteins. Compounds that modulate the prostate specific or testis
 CC specific polypeptide are useful to diagnose, prevent or treat disorders
 CC of the testis or prostate particularly prostate cancer, benign
 CC prostatic hyperplasia, acute prostatitis, testicular cancer,
 CC cryptorchidism, undescended, retractile, ascending or vanished
 CC testis. Other proliferative disorders for which the modulators may be
 CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
 CC cancer, pancreatic cancer, liver cancer and lung cancer. The
 CC present sequence represents a prostate specific protein, Six-
 CC Transmembrane Protein of Prostate 1, STM1.
 CC
 XX
 SQ Sequence 490 AA;
 Query Match 99.8%; Score 2530; DB 22; Length 490;
 Best Local Similarity 99.8%; Pred. No. 4.8e-259;
 Matches 489; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MESISMGSPKSLSECLPNCINGIKDARKVTGVIGSGDFAKSLTRLRCGYHVVIGS 60
 DB 1 mesismgspkslsetclpncingikdarkvtgvigsgdfaksltrlrcgyhvvigs 60
 QY 61 RNPFASEFPFHVVDYHHEDAFKTNIRFAHREHTSLMDLRHLVGLIDVSNM 120
 DB 61 rnpfasefphvvdvthbedalktnlrfahrehtslmdlrhlvlgkllidvsnm 120
 QY 121 RINQYPSNNAEYLALFPDLSLVKGFNVVSAMALQSGPKDASROYICSNNOAROOYIE 180
 DB 121 ringypsnaeylalfpdsllvkgfnvvsamalgpkdasryylicsnnidargyie 180
 QY 181 LARQNLNPIIDGLSSASAREIENLPLRLFTLMRGPVVVAISLAFEFYFVRDVIHPYA 240
 DB 181 larqnlndpiddglssasareienplrlftlmgpvvvaalslafeflyfvrdivhpya 240
 QY 241 RNOQSDRYKPIRIEYVNTPLPVATITLSLVYLAGLAAAYLYGTRRRPPELEWLO 300
 DB 241 rnoqsdrykpielievntplpvaititlslylaglaaaylygtrrrpplwlewlq 300
 QY 301 CRKQLGLISFEFAWVAVSYLCLPMRSEERYLFNMAVYOOVHANINSMNEEYWRITEMY 360
 DB 301 crkqlglisffefawvavayclpmrserylfnmayyqvahansmneewrilemy 360
 QY 361 ISFGINSLGLISLAVTSIDSVSNALNMREESFIQSTLGVYALLISFHVLIYGMKRAFE 420
 DB 361 isfginlglisllavtsidsvsnalnwmreesfiqstlgyvallisfhnvliywkrafe 420
 QY 421 EEYRFTYPRNFVALVLPSTVILGKITPLPCISRKLRKRWKWEWSQPLEBIGTIP 480
 DB 421 eeyrftypnfvalvlpstvilgkittlpcisrklrkrrkwewsqplebigtipp 480

DB 421 eeyrftypnfvalvlpstvilgkittlpcisrklrkrrkwewsqfleeglgctip 480
 QY 481 HVSPERYTVM 490
 DB 481 hvspertyvm 490
 RESULT 2
 AAU10188
 ID AAU10188 standard; Protein; 454 AA.
 XX
 AC AAU10188;
 XX
 XX 16-JAN-2002 (first entry)
 DE Human ORF2 of Six-Transmembrane Protein of Prostate 1.
 XX
 KW Human; Six-Transmembrane Protein of Prostate 1; STM1; prostate cancer;
 KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
 KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
 KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
 KW liver cancer; lung cancer; cytostatic; ORF2.
 XX
 OS Homo sapiens.
 XX
 XX WO200172962-A2.
 XX
 XX 04-OCT-2001.
 XX
 XX 23-MAR-2001; 2001WO-US09410.
 XX
 XX 24-MAR-2000; 2000US-191929P.
 XX
 PA (SAAT/) SAATC1OGLU F.
 XX
 PI Saatioglu F.
 XX
 XX WPI: 2001-662926/76.
 DR N-PSDB; AAS15810, AAS15811.
 XX
 PT New polynucleotide for the diagnosis, prevention and treatment for
 PT prostate and testis disorders, particularly prostate cancer, comprises
 PT prostate-specific or testis-specific nucleic acids
 XX
 PS Claim 1; Fig 4H; 114pp; English.
 XX
 CC The invention relates to substantially pure prostate-specific or
 CC testis-specific polypeptides and the nucleic acids encoding them.
 CC Also included are vectors and host cells expressing the proteins, a
 CC transgenic animal expressing the protein, antibodies against the
 CC proteins, probes for detecting the nucleic acids, antisense molecules
 CC for the nucleic acids and methods of isolating modulators of the
 CC proteins. Compounds that modulate the prostate specific or testis
 CC specific polypeptide are useful to diagnose, prevent or treat disorders
 CC of the testis or prostate particularly prostate cancer, benign
 CC prostatic hyperplasia, acute prostatitis, testicular cancer,
 CC cryptorchidism, undescended, retractile, ascending or vanished
 CC testis. Other proliferative disorders for which the modulators may be
 CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
 CC cancer, pancreatic cancer, liver cancer and lung cancer. The
 CC present sequence is prostate specific protein, Six-Transmembrane
 CC Protein of Prostate 1, STM1, ORF2.
 CC
 XX
 SQ Sequence 454 AA;
 Query Match 90.5%; Score 2294; DB 22; Length 454;
 Best Local Similarity 100.0%; Pred. No. 4.4e-234;
 Matches 444; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MESISMGSPKSLSECLPNCINGIKDARKVTGVIGSGDFAKSLTRLRCGYHVVIGS 60
 DB 1 mesismgspkslsetclpncingikdarkvtgvigsgdfaksltrlrcgyhvvigs 60

```
OY 61 RNPFASEFPYVDVTHHEDALTKTNIEFVAIHREHYTSLMDLRHLVGLKILIDVSNMM 120
    |||||||
DB 61 rnpkfasefpyvvdvthheda|tktni|l|fva|ihrehyts|lwdl|rl|lygk|il|dvsnm| 120
OY 121 RINQYPSNAEYLASLEPDSLIVKGFNVYSAMALQCPKASROYICSNNIQARQOVI 180
    |||||||
DB 121 rinqypsnaey|laas|lfps|llv|kgfnvsawa|l|q|p|das|q|y|ic|sn|ni|q|ar|q|v|le 180
OY 181 LARQLNFRIDIGSLSSAENEIENLPLRTLTMRGPVVAISLATFEFLYSFVRDVIHPYA 240
    |||||||
DB 181 larq|ln|fr|id|ig|sl|ssa|ene|en|l|p|rl|tl|w|rg|pv|va|is|la|t|fe|fl|ys|f|vr|dv|ih|pya 240
OY 241 RNOQSDFYKPIEIVNKTPIVAITLSTLYVLAGLAAAYOYYGKRRPFWLTWQ 300
    |||||||
DB 241 rnoqsd|fy|k|pi|e|iv|n|kt|p|i|va|it|ls|t|ly|v|la|g|l|aa|ay|q|y|y|g|k|rr|f|w|l|t|w|q 300
OY 301 CRKQGLSFPFAMVAVSLCLPMRBSRYFLNMAVOOVANIENTSNNEEVRRIEMX 360
    |||||||
DB 301 crk|q|g|l|s|f|p|f|a|mv|av|s|l|c|l|p|mr|bs|ry|f|l|n|ma|vo|ov|an|ien|ts|n|ne|e|vr|ri|em|x 360
OY 361 ISFGIMSLGLSLAVTSIPSVSNALNMRREFSTQSTGLYVALLSTFHVLIYWKRAFE 420
    |||||||
DB 361 isf|g|im|s|l|g|l|s|l|av|ts|ip|sv|sn|al|n|mr|ref|st|q|st|g|l|y|v|al|l|st|f|hv|li|y|w|k|ra|f|e 420
OY 421 EEEYRFPNPFVIALVPSIVIL 444
    |||||||
DB 421 eey|y|r|f|p|n|p|f|v|ia|l|v|p|s|i|v|i|l 444
```

RESULT 3
AAE02781 ID AAE02781 standard: Protein; 454 AA.

XX AAE02781;

XX 06-AUG-2001 (first entry)

XX Human six transmembrane epithelial antigen of prostate (STEAP)-2 protein.

XX Human; cytosolic; antiproliferative; vaccine; gene therapy;

XX six transmembrane epithelial antigen of the prostate-2; STEAP-2;

XX chromosome 7q21; cancer; prostate; colon; bladder; lung; ovarian;

XX pancreatic.

XX Homo sapiens.

XX Key

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

XX Region

Location/Qualifiers
100..108
/label= HLA-A2-binding-peptide #5
153..165
/label= Immunogenic-peptide #1
207..228
/label= Transmembrane_domain #1
227..235
/label= HLA-A2-binding-peptide #1
255..277
/label= Transmembrane_domain #2
304..325
/label= Transmembrane_domain #3
306..314
/label= HLA-A2-binding-peptide #4
307..315
/label= HLA-A2-binding-peptide #3
345..358
/label= Immunogenic-peptide #2
359..381
/label= Transmembrane_domain #4
393..415
/label= Transmembrane_domain #5
402..410
/label= HLA-A2-binding-peptide #2
428..450

FT XX /label= Transmembrane_domain #6

PN XX WO200140276-A2.

PD XX 07-JUN-2001.

PF XX 06-DEC-2000; 2000WO-US33040.

PR XX 06-DEC-1999; 99US-0455486.

PA XX (UROG-) UROGENESYS INC.

PI PI Afar DEH, Hubert RS, Raitano AB, Safran DC, Mitchell SC, Faris M;

DR DR Jakobovits A;

DR DR WPI: 2001-367804/38.

DR DR N-PSDB; AAD07072.

PT PT New STEAP (six transmembrane epithelial antigen of the prostate)

PT PT proteins, expressed in human cancers, useful for detecting and treating

PS PS cancer -

PS PS Claim 1; Fig 9A-9C; 187pp; English.

CC The present sequence is human six transmembrane epithelial antigen of

CC the prostate (STEAP)-2 protein. STEAP is a member of cell

CC surface serpentine transmembrane antigens. STEAP-2 gene is located on

CC chromosome 7q21 and is used in gene therapy. Inhibiting the development

CC or progression of a cancer (eg. prostate, colon, bladder, lung, ovarian

CC and pancreatic) expressing STEAP or inhibiting growth or killing cells

CC expressing STEAP in a patient, comprises administering a vaccine

CC composition to the patient. Treating a patient with a cancer that

CC expresses STEAP, or inhibiting growth or killing cells expressing STEAP,

CC comprises administering to the patient a vector encoding single chain

CC monoclonal antibody that comprises the variable domains of the heavy and

CC light chains of the monoclonal antibody that specifically binds to STEAP,

CC such that the vector delivers the single chain monoclonal antibody coding

CC sequence to the cancer cells and the encoded single chain monoclonal

CC antibody is expressed intracellularly.

CC Note: This sequence is stated to be the same as that being shown as

CC SEQ ID NO:8 (AAE02841) in sequence listing of the specification. However

CC both the sequences differ at several positions.

XX Sequence 454 AA:

SO Query Match 90.5%; Score 2294; DB 22; Length 454;
Best Local Similarity 100.0%; Pred. No. 4.4e-234;
Matches 444; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MESISMGSPKSLSECTCLNGINGIRIDARKVYGVYIGSGDFAKSLTRILRCGYHYVIGS 60

DB 1 mesismmgspk|sl|se|ct|cl|ng|ing|ir|id|ark|vy|gv|y|ig|sg|df|ak|sl|tr|il|rc|gy|hy|v|ig|s 60

OY 61 RNPFASEFPYVDVTHHEDALTKTNIEFVAIHREHYTSLMDLRHLVGLKILIDVSNMM 120

DB 61 rnpkfasefpyvvdvthheda|tktni|l|fva|ihrehyts|lwdl|rl|lygk|il|dvsnm| 120

OY 121 RINQYPSNAEYLASLEPDSLIVKGFNVYSAMALQCPKASROYICSNNIQARQOVI 180

DB 121 rinqypsnaey|laas|lfps|llv|kgfnvsawa|l|q|p|das|q|y|ic|sn|ni|q|ar|q|v|le 180

OY 181 LARQLNFRIDIGSLSSAENEIENLPLRTLTMRGPVVAISLATFEFLYSFVRDVIHPYA 240

DB 181 larq|ln|fr|id|ig|sl|ssa|ene|en|l|p|rl|tl|w|rg|pv|va|is|la|t|fe|fl|ys|f|vr|dv|ih|pya 240

OY 241 RNOQSDFYKPIEIVNKTPIVAITLSTLYVLAGLAAAYOYYGKRRPFWLTWQ 300

DB 241 rnoqsd|fy|k|pi|e|iv|n|kt|p|i|va|it|ls|t|ly|v|la|g|l|aa|ay|q|y|y|g|k|rr|f|w|l|t|w|q 300

OY 301 CRKQGLSFPFAMVAVSLCLPMRBSRYFLNMAVOOVANIENTSNNEEVRRIEMX 360

DB 301 crk|q|g|l|s|f|p|f|a|mv|av|s|l|c|l|p|mr|bs|ry|f|l|n|ma|vo|ov|an|ien|ts|n|ne|e|vr|ri|em|x 360

QY 361 ISFGIMSLGLSLAVTSIPSVSNALMWREFSTQSTLGYVALLISFPHVLLYGKRAFE 420
|||
Db 361 ISFGIMSLGLSLAVTSIPSVSNALMWREFSTQSTLGYVALLISFPHVLLYGKRAFE 420
|||
QY 421 EEFYRFTTPNFVVALVPSIVIL 444
|||
Db 421 eeyrfttppnfvalvpsivil 444
|||
RESULT 4
ID AAU10189 standard; Protein: 419 AA.
XX
AC AAU10189;
XX
DE 16-JAN-2002 (first entry)
XX
DE Human ORF3 of Six-Transmembrane Protein of Prostate 1.
XX
XX Human; Six-Transmembrane Protein of Prostate 1; STM1; prostate cancer;
KM benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KM cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KM leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KM liver cancer; lung cancer; cytostatic; ORF3.
XX
OS Homo sapiens.
XX
PN WO200172962-A2.
XX
PD 04-OCT-2001.
XX
PF 23-MAR-2001; 2001MO-US09410.
XX
PR 24-MAR-2000; 2000US-191929P.
XX
XX (SAAT/) SAATCTOGLU F.
PA
PI Saactioglu F;
PI
DR WPI: 2001-662926/76.
DR N-PSDB; AAS15813.
XX
XX New polynucleotide for the diagnosis, prevention and treatment for
PT prostate and testis disorders, particularly prostate cancer, comprises
PT prostate-specific or testis-specific nucleic acids -
XX
XX Claim 1; Fig 4K; 114pp; English.
PS
XX The invention relates to substantially pure prostate-specific or
CC testis-specific polypeptides and the nucleic acids encoding them.
CC Also included are vectors and host cells expressing the proteins, a
CC transgenic animal expressing the protein, antibodies against the
CC proteins, probes for detecting the nucleic acids, antisense molecules
CC for the nucleic acids and methods of isolating modulators of the
CC proteins. Compounds that modulate the prostate specific or testis
CC specific polypeptide are useful to diagnose, prevent or treat disorders
CC of the testis or prostate particularly prostate cancer, benign
CC prostatic hyperplasia, acute prostatitis, testicular cancer,
CC cryptorchidism, undescended, retractile, ascending or vanished
CC testis. Other proliferative disorders for which the modulators may be
CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC cancer, pancreatic cancer, liver cancer and lung cancer. The
CC present sequence is prostate specific protein, Six-Transmembrane
CC Protein of Prostate 1, STM1, ORF3.
CC
XX
XX Sequence 419 AA;

Query Match 80.3%; Score 2036; DB 22; Length 419;
Best Local Similarity 99.7%; Pred. No. 8,6e-207;
Matches 394; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MESISMGSPKSLSEETCLPNCINGIKDARKVTGVIIGSGDFAKSLTIRLRCGYHVVIGS 60
|||
Db 1 mesismgspkslsetcnpingikdarkvtgvisgdfaksltirlrcgyhvvigs 60
|||
QY 61 RNPFASEFPFHVVDYTHHEALTKNTNIFAIHREHTSMDLRHLLYCKILLDVSNNM 120
|||
Db 61 rnpfasefpfhvvdvthhedalkntnifaihrehytsmdlrhllyckilldvsnnm 120
|||
QY 121 RINQYPSNAEYSLAPFDSLIVKGFNVVSAMALQLPKDSROVYICSNNIQARQOYIE 180
|||
Db 121 ringypsnaeyslapfdsliivkgfnvsawalqpkdsroryicsnniqarqoyie 180
|||
QY 181 LARQLNFPIDLGSLSSAREIENLPLRLFTLMRGPVVAISLATFEFLSVFROVHHPYA 240
|||
Db 181 larglnfpidlgslssareienlplrlftlmgpvvaistlatfefflsvfrdvhpya 240
|||
QY 241 RNOQSDPKRPIREIVNKTLPVATITLSIVYLAQLLAQVLYGTYRRPPLLETWLO 300
|||
Db 241 rnoqsdpykripieivnktlpvatitlsivylaqlllaavlygtyrrpppletwlo 300
|||
QY 301 CRKOLGSLFFFAVAVHVAYSLSCLPMRSEERYLFNMAVQOVHANIENSMNEEYVRIEMY 360
|||
Db 301 crkqlgslsfffavavhvayslclpmrserylfnmayqovhansiensneeeyvriemy 360
|||
QY 361 ISFGIMSLGLSLAVTSIPSVSNALMWREFSTQ 395
|||
Db 361 ISFGIMSLGLSLAVTSIPSVSNALMWREFSTQ 395
|||

RESULT 5

ABG12306 standard; Protein: 576 AA.

ABG12306;

18-FEB-2002 (first entry)

Novel human diagnostic protein #12297.

Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder.

Homo sapiens.

WO200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001MO-US08631.

31-MAR-2000; 2000US-0540217.

23-AUG-2000; 2000US-0649167.

(HYSE-) HYSEQ INC.

Drmanac RT, Liu C, Tang YT;

WPI: 2001-639362/73.

N-PSDB; AAS76493.

New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

Claim 20; SEQ ID NO 42665; 103pp; English.

The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (II) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG3037 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 576 AA:

Query Match 77.4%; Score 1962.5; DB 22; Length 576;
Best Local Similarity 87.4%; Pred. No. 8.6e-199;

Matches 388; Conservative 0; Mismatches 1; Indels 55; Gaps 1;

QY 1 MESISMGSPKLSLSETCLEPENGINGIDARKVTVGVIGSGDFAKSLTRLRGCVHVVIGS 60
DB 1 mesismgspkslsetflpnngingidarkvkvvgisgdfaksltrlrcgyhvvgis 60
QY 61 RNKFASEFPFHVVDVTHHEDALTKNTNIFVAHREHYTSLMDRLHLVKGKILIDVSNM 120
DB 61 rnkfasetfpfvvdvthhedaltknltfvalhrehytslwdrlhlvkgkildvsnm 120
QY 121 RINQYPSNAEYLASEPDSLIVKGFNVVSAMALQGPDKASROYVICSNNIQARQOYIE 180
DB 121 rinqypsnaeylaslfpdslivkgfnvvsawaldqgpkdasroyvicsnniqarqvie 180
QY 181 LARQLNFIPIIDGLSSASAREIENLPRLFTLMRGPPVVAISLATEFFLYSFVADVHPYA 240
DB 181 larqlnfiipidglssasareienlprlftlmrgppvvaistlatefflysfvadvhpya 240
QY 241 RNOQSDFYKPIEIVNKTPIVAITLSLYLAGLAAAYQLYGKRYRFPMTWTWIO 300
DB 241 rnoqsdfykipieivnktlpivaitlslvylagllaaayqlygkryrrfpwletwlg 300
QY 301 CRKOLGLSFFPAMVAVAYSLCLPMRSEERYFLNMAVOOVHANINENSNNEEVRRIEMV 360
DB 301 crkoglslsffpamvavayslclpmrserylflnmayq----- 339
QY 361 ISFGIMSLGLSLAVTSIPSVSNALNMRFEFSTIGVYALLISTFHVLYIGMKRAFE 420
DB 361 isfgimslglslavtsipsvsnalnmrfeefstigvyalistfhvlyigmkrate 420
QY 421 EEEYRFTPPNRYLALVPSIYIL 444
DB 421 eeyrftppnrylavlpsiyil 389

RESULT 6

ID ABG00113 standard; Protein; 1273 AA.

AC ABG00113;

DT 13-FEB-2002 (first entry)

XX Novel human diagnostic protein #104.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

XX WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

XX N-PSDB; AAS64300.

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

XX Claim 20; SEQ ID No 30472; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

XX polypeptide (II) sequences. (I) is useful as hybridisation probes,

XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome

XX and gene mapping, and in recombinant production of (II). The

XX polynucleotides are also used in diagnostics as expressed sequence tags

XX for identifying expressed genes. (I) is useful in gene therapy techniques

XX to restore normal activity of (II) or to treat disease states involving

XX (II). (II) is useful for generating antibodies against it, detecting or

XX quantitating a polypeptide in tissue, as molecular weight markers and as

XX a food supplement. (II) and its binding partners are useful in medical

XX imaging of sites expressing (II). (I) and (II) are useful for treating

XX disorders involving aberrant protein expression or biological activity.

XX The polypeptide and polynucleotide sequences have applications in

XX diagnostics, forensics, gene mapping, identification of mutations

XX responsible for genetic disorders or other traits to assess biodiversity

XX and to produce other types of data and products dependent on DNA and

XX amino acid sequences. ABG00010-ABG3037 represent novel human

XX diagnostic amino acid sequences of the invention.

XX Note: The sequence data for this patent did not appear in the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1273 AA;

Query Match 77.4%; Score 1962.5; DB 22; Length 1273;

Best Local Similarity 87.4%; Pred. No. 2.9e-198;

Matches 388; Conservative 0; Mismatches 1; Indels 55; Gaps 1;

QY 1 MESISMGSPKLSLSETCLEPENGINGIDARKVTVGVIGSGDFAKSLTRLRGCVHVVIGS 60
DB 1 mesismgspkslsetflpnngingidarkvkvvgisgdfaksltrlrcgyhvvgis 60
QY 61 RNKFASEFPFHVVDVTHHEDALTKNTNIFVAHREHYTSLMDRLHLVKGKILIDVSNM 120
DB 61 rnkfasetfpfvvdvthhedaltknltfvalhrehytslwdrlhlvkgkildvsnm 120
QY 121 RINQYPSNAEYLASEPDSLIVKGFNVVSAMALQGPDKASROYVICSNNIQARQOYIE 180
DB 121 rinqypsnaeylaslfpdslivkgfnvvsawaldqgpkdasroyvicsnniqarqvie 180
QY 181 LARQLNFIPIIDGLSSASAREIENLPRLFTLMRGPPVVAISLATEFFLYSFVADVHPYA 240
DB 181 larqlnfiipidglssasareienlprlftlmrgppvvaistlatefflysfvadvhpya 240
QY 241 RNOQSDFYKPIEIVNKTPIVAITLSLYLAGLAAAYQLYGKRYRFPMTWTWIO 300
DB 241 rnoqsdfykipieivnktlpivaitlslvylagllaaayqlygkryrrfpwletwlg 300
QY 301 CRKOLGLSFFPAMVAVAYSLCLPMRSEERYFLNMAVOOVHANINENSNNEEVRRIEMV 360
DB 301 crkoglslsffpamvavayslclpmrserylflnmayq----- 411

QY 361 ISFIMSLGLSLAVTSIPSVNALMWRFSFIQSTGLVALLSTFHVLIYGMKRAFE 420
 DB 412 -----gstlgyvalliststhlvlygmkrate 437
 QY 421 EEYRFTYPPNFVLAIVPSIVIL 444
 DB 438 eeyrftypnfvlalvpsivil 461
 RESULT 7
 ID AAE02841 standard; protein: 450 AA.
 AC AAE02841;
 DT 06-AUG-2001 (first entry)
 DE Human STEAP-2 protein, alternative version.
 XX
 KM Human: cytostatic; antiproliferative; vaccine; gene therapy;
 KM six transmembrane epithelial antigen of the prostate-2; STEAP-2;
 KM chromosome 7q21; cancer; prostate; colon; bladder; lung; ovarian;
 KM pancreatic.
 XX
 OS Homo sapiens.
 XX
 FH Key location/Qualifiers
 FT Misc-difference 335..336
 FT Misc-difference 375..376 /note="Encoded by GACTGAGCT"
 FT Misc-difference 415..416 /note="Encoded by AGATGAAGT"
 FT Misc-difference 445..446 /note="Encoded by GAGTAAGC"
 FT Misc-difference 448..449 /note="Encoded by ACATGAAGT"
 FT Misc-difference 448..449 /note="Encoded by AATTAAATTC"
 XX
 PN WO200140276-A2.
 PD 07-JUN-2001.
 XX
 PF 06-DEC-2000; 2000WO-US33040.
 XX
 PR 06-DEC-1999; 99US-0455486.
 XX
 PA (UROG-) UROGENESYS INC.
 PI Afar DEH, Hubert RS, Raitano AB, Saffran DC, Mitchell SC, Paris M;
 PI Jakobovits A;
 XX
 DR WPI: 2001-367804/38.
 DR N-PSDB; AAD07072.
 XX
 PT New STEAP (six transmembrane epithelial antigen of the prostate)
 PT proteins, expressed in human cancers, useful for detecting and treating
 PT cancer.
 XX
 PS Claim 1; Page 175-176; 187pp; English.
 XX
 CC The present sequence is an alternative version of human six transmembrane
 CC epithelial antigen of the prostate (STEAP)-2 protein. STEAP is a member
 CC of cell surface serpentine transmembrane antigens. STEAP-2 gene is
 CC located on chromosome 7q21 and is used in gene therapy. Inhibiting the
 CC development or progression of a cancer (eg. prostate, colon, bladder,
 CC lung, ovarian and pancreatic) expressing STEAP or inhibiting growth or
 CC killing cells expressing STEAP in a patient, comprises administering a
 CC vaccine composition to the patient. Treating a patient with a cancer that
 CC expresses STEAP, or inhibiting growth or killing cells expressing STEAP,
 CC comprises administering to the patient a vector encoding single chain
 CC monoclonal antibody that comprises the variable domains of the heavy and
 CC light chains of the monoclonal antibody that specifically binds to STEAP.

CC such that the vector delivers the single chain monoclonal antibody coding
 CC sequence to the cancer cells and the encoded single chain monoclonal
 CC antibody is expressed intracellularly.
 CC Note: This sequence is stated to be the same as that being shown as
 CC SEQ ID NO:8 (AAE02781) in figure 9A-9C of the specification. However
 CC the present sequence lacks several amino acids at its N-terminal end
 CC and has additional amino acids at its C-terminal end when compared with
 CC the sequence shown in figure 9A-9C.
 XX
 SQ Sequence 450 AA;
 Query Match 66.4%; Score 1683; DB 22; Length 450;
 Best Local Similarity 94.6%; Pred. No. 2.5e-169;
 Matches 332; Conservative 1; Mismatches 8; Indels 10; Gaps 1;
 QY 119 NMRINQPSNAEYLAIFPDSILYKGVNYSANALQGPDAHQYICNNIQANQY 178
 DB 1 nmrlngqpsnaeylasifpdsilvkgfnvsawalqlpksarqyicnnlqarqv 60
 QY 179 IELAROLNFIPIIDIGSLSSAREIENLPLRLFTLMRGPVVAISLATEFFELXSFVRYIHP 238
 DB 61 Ielarlfnlfpidlgslssareienlplrlftlwrpvaaislatffllysvrdvlnp 120
 QY 239 YARNQSDPYKPIPIEYVKTLPIVATLLSLVYLAGLAAAYQLYGTXYRRPPELWTW 298
 DB 121 yarngsdpykpiplieynktlplivatllslvylagllaaayqlgtkyrrfpplwtw 180
 QY 299 LQCRKQGLSFFRPMHVAISLCLPQRSRYFLFMNAQOVNANINENSGNEEVRIE 358
 DB 181 lqcrkqglslsffamhvaslclpqrserylflmayqqvnaansvneeevrlw 240
 QY 359 MYISFGIMSLGLSLAVTSIPSVNALMWRFSFIQSTGLVALLSTFHVLIYGMKRA 418
 DB 241 myisfgimslglslavtsipsvsnalnwrefisgltyvallstfhlvlygmkra 300
 QY 419 FEEERYRTPPNFVLAIVPSIVIL-----FLPCISRKRL 459
 DB 301 feeeryrftypnfvlalvpsivilqlcrypdagtlcldstslk 351
 RESULT 8
 ID AAB49483 standard; protein: 488 AA.
 AC AAB49483;
 DT 08-MAR-2001 (first entry)
 DE Rat p-HYDE.
 XX
 KM Rat: p-HYDE: cytostatic; gene therapy; apoptosis; leukemia; prostate;
 KM tumour suppressor gene; DNA repair; cancer; melanoma; lymphoma;
 KM colorectal; pancreatic; breast; brain; gastric carcinoma.
 XX
 OS Rattus sp.
 XX
 PN WO200071564-A2.
 PD 30-NOV-2000.
 XX
 PF 01-MAY-2000; 2000WO-US11456.
 XX
 PR 29-APR-1999; 99US-0131607.
 PR 29-APR-1999; 99US-0302457.
 PR 26-NOV-1999; 99US-0499817.
 XX
 PA (UYTE-) UNIV TENNESSEE RES CORP.
 PI Steiner MS, Wang C, Rinaldy A, Menon R;
 XX
 DR WPI: 2001-032016/04.
 DR N-PSDB; AAC89167.

DR N-PSDB; AAS08646.

XX New G-protein-coupled receptor-like polypeptides and polynucleotides,
PT useful for treating diseases of ophthalmic, neurological, immunological
PT and nephritic systems and hormonal dysfunction, cancer, atherosclerosis
PT and diabetes -

XX Claim 10; Page 164-166; 259pp; English.

XX The sequence represents a human G-protein coupled receptor (GPCR)-like
CC protein, found to have sequence similarity to human STRAP-1 protein.
CC The GPCR-like polypeptides and polynucleotides are useful for
CC the treatment of diseases of ophthalmic, neurological (e.g. Alzheimer's
CC disease and Parkinson's disease, immunological (e.g. HIV infection and
CC candidiasis), autoimmune disorders (e.g. multiple sclerosis, systemic
CC lupus erythematosus and rheumatoid arthritis), platelet disorders (e.g.
CC thrombocytopenia and aplastic anaemia), inflammatory disorders (e.g.
CC septic shock and systemic inflammatory response syndrome, SIRS) and
CC nephritic systems. They may also be used to treat hormonal dysfunction,
CC cancer, atherosclerosis, wounds, tissue regeneration, haemophilia,
CC leukaemia, reperfusion injury, psoriasis and diabetes. Numerous examples
CC of each type of disorder are given in the specification. Anti-GPCR-like
CC protein antibodies are useful for detecting or quantitating the
CC polypeptide in tissue. The polypeptides can also be used as molecular
CC weight markers and as a food supplement.

XX Sequence 488 AA:

Query Match 53.5%; Score 1356; DB 22; Length 488;

Best Local Similarity 53.1%; Pred. No. 1.3e-134;
Matches 260; Conservative 91; Mismatches 121; Indels 18; Gaps 6;

OY 3 SISMGSPKSLSEETCLPNCINGIKDARKVTYVIGSGDFAKSLTIRLCGHHVIGSRN 62
DB 11 slhlydsdsjak--vpd-----eapK--vgllsgdfatrlatrlvsgtkvvvgsrn 60
OY 63 PFASEFPFHVVDVTHHEDALTKTNIIFVAIHREHYTSLMDRLHLLVGCILIDVSN--NM 120
DB 61 pKtarlrlpsaaqvtfqgeaavspevlfavtrehytslcsldqkglldvdsnpieq 120
OY 121 RINQYSPESNAEYLASLPDLSLVKGFNVVSAWALQGPDKASROYVYSNNIQARQVTE 180
DB 121 ehlgtrresnaeylaalfpctcvkafnvysavtlqagprdgrrpylcsqdgqdeakraxe 180
OY 181 LARQNLFPIDGSSRSAREIENLELFTLMKRGVVAISLATPFLYSEPRDVIHRA 240
DB 181 malangfmpvdmgslasaweampdlrlpawkvptllaqlfvcfyaynftvrdvlpqv 240
OY 241 RNOQSDFFKPIPIELVNTLPIVAITLISLVYLAGLLAAVQLYGTGRFRPMELETMQ 300
DB 241 qesgqkffllpsvvnltlpcvayllstvylypvaalqrlrrtkygrfvdwldwlg 300
OY 301 CRKQGLSLFFPAMVAVYSLCLPMRSEERYELFMAYQOVHANIENTSNVEEVRIMEX 360
DB 301 htkqigllstfcaahalsfclprtrahrydlvnlavkqvalankshllweeevrmeyi 360
OY 361 ISFGIMSLGLSLAVTSPYSNLMNREPSIOSTIGYVALLSTFHVILYGGKRAFE 420
DB 361 lslgyalagtlslavtslpslanslwnreitsfvgsslgfvalvstlhtllygwrlde 420
OY 421 EBYVRFYPPNPFVLAFLPSVILKILFLPCISRKLRKIKKGWKSQFLEGGIGGTRP 480
DB 421 eerykyfypfttlcllpvcvllakalflipclstlrlarlrgrwer-----estlkfclp 476
OY 481 --HVSPEVYT 488
DB 477 tchalaekts 486

RESULT 12

AAB93224 standard; Protein; 488 AA.

XX AAB93224;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:12212.

XX Human; primer: detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX Ep1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Claim 8; SEQ ID 12212; 2537Pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification, where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 488 AA:

Query Match 53.5%; Score 1356; DB 22; Length 488;

Best Local Similarity 53.1%; Pred. No. 1.3e-134;
Matches 260; Conservative 91; Mismatches 121; Indels 18; Gaps 6;

OY 3 SISMGSPKSLSEETCLPNCINGIKDARKVTYVIGSGDFAKSLTIRLCGHHVIGSRN 62
DB 11 slhlydsdsjak--vpd-----eapK--vgllsgdfatrlatrlvsgtkvvvgsrn 60
OY 63 PFASEFPFHVVDVTHHEDALTKTNIIFVAIHREHYTSLMDRLHLLVGCILIDVSN--NM 120
DB 61 pKtarlrlpsaaqvtfqgeaavspevlfavtrehytslcsldqkglldvdsnpieq 120

RESULT 2
AC2560
hypothetical protein alr8074 [imported] - Anabaena sp. (strain PCC 7120) plasmid pcc7120
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AC2560
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2560
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-211 <KUR>
A:Cross-references: GB:AP003603; PIDN:BA877404.1; PID:g17134847; GSPDB:GN00182
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr8074
A:Genome: plasmid

Query Match 6.3%; Score 159.5; DB 2; Length 211;
Best Local Similarity 24.9%; Pred. No. 3.1e-05;
Matches 51; Conservative 43; Mismatches 70; Indels 41; Gaps 8;
QY 33 VGVIGSGDFAKSLIRLCRCGVHVVIGSRNP-----KFASEFPFHVDVTHHEDALTKTN 87
DB 3 IAFIGIGVGSALASQLSDHVTIARNSNSDVKALKKYP-LOVSPQEAIAQAE 61
QY 88 IIFVA-----IHRHYTSLMDRLHLVGLKILIDVSNMNRIN-----QYPSNAEYLASLP 138
DB 62 VIFLPTPTTAQALAEVGDLS---GKILVDTNPVGNVTHGLKBSQSEIIVOSFVP 117
QY 139 DSLIVKGFNVVSANALQGPDKASROY-----ICSNNIOAQOYIELARQINF 187
DB 118 HAKVAKATTI-----YGFNFENHTYPGYGNLKPAMLIAGNDVPARQVVTLCQQLGW 170
QY 188 IPIDGLSSAREIENTPLRLFTLM 212
DB 171 EAVDVGNLSMLHLEHMTL-----LM 191
RESULT 3
T50571
probable oxidoreductase [imported] - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 28-Jul-2000
C:Accession: T50571
R:Redenbach, M.; Kleiser, H.M.; Denapalte, D.; Eichner, A.; Cullum, J.; Kinashl, H.; Hopa, M.
Mol. Microbiol. 21, 77-96, 1996
A:Title: A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb S.
A:Reference number: Z20556; MUID:97000351
A:Accession: T50571
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-239 <RED>
A:Cross-references: EMBL:AL133220; PIDN:CAB61708.1
C:Genetics:
A:Experimental source: strain A3(2)
A:Note: SCC75A.08c
C:Superfamily: conserved hypothetical protein MJ1501

Query Match 6.2%; Score 156.5; DB 2; Length 239;
Best Local Similarity 27.9%; Pred. No. 6.2e-05;
Matches 61; Conservative 41; Mismatches 72; Indels 45; Gaps 10;
QY 25 IKDKRKYTVGVI-GSGDFAKSLIRLCRCGVHVVIGSR---NPKFASEFPFHVDVTHHNE 80
DB 23 LPDVSGLVGVGTFQGGKGLAVRLAKAGKQVIVGSAERAAAEIGHGVEGADNA 82

QY 81 DALTKNTIIFVAIHRE-HYTSIMDLRLHLVGLKILIDVSNMNRINO-----YPE--SNAE 131
DB 83 ETARSDVVIYAVPMDGKLTLESIRALSGKIVDVCNPLGFDKAGYALKPEEGSAAE 142
QY 132 YLASFPDLSLVKGFNVVSANALQ-----LGPDKASROYICSNNTIOARQV 178
DB 143 OAAALLPDSRAAAAFHLSAVLLQDPEIDETDVMVAGEERADVEI-----VOA----- 192
QY 179 IEIARQLNFIPI---IDGLSSAREIENTPLRLFTLMR 213
DB 193 --LAGR---IFGMRCVFAGRLRNHMQVESIVANLISVNR 226

RESULT 4
H69400
conserved hypothetical protein AF1209 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999
C:Accession: H69400
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod, G.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes, Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing, arch.
A:Reference number: A69250; MUID:98049343
A:Accession: H69400
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-213 <KLE>
A:Cross-references: GB:AE001021; GB:AE00782; MID:92669344; PIDN:AA890038.1; PID:9264
C:Superfamily: conserved hypothetical protein MJ1501

Query Match 5.8%; Score 147; DB 2; Length 213;
Best Local Similarity 27.5%; Pred. No. 0.0003;
Matches 56; Conservative 44; Mismatches 68; Indels 36; Gaps 11;
QY 37 GSGDFAKSLIRLCRCGVHVVIGSRNP-----KFASEFPFHVDVTHHEDALTKTN 87
DB 8 GTGNIGEGELARWKGKLTGIVTSGSKLEAKLASDYIKKAGDASIIIGKREDAAETCD 66
QY 89 IFVAIHREHYTSLMD-----LRHLVGLKILID-----VSNMNRINOPE--SNAE 131
DB 67 --VAV---FTIPWFAFDTAEMLRQLAGKVIISLPVPMKVDNF--VYVRPEGSAAE 119
QY 132 YLASFPDLSLVKGFNVVSANALQGPDKASROYICSNNTIOARQOYIELARQINF-PI 190
DB 120 KLASVLEESSVVAAYHSHIPARRFANLGEFEMDVPICGDS-GAREVYVDLTKISGLRAL 178
QY 191 DLGSLSSAREIENT-PLRLFTLMR 213
DB 179 DAGLSNAHLVESLTPILINMKR 202

RESULT 5
D69361
conserved hypothetical protein AF0892 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999
C:Accession: D69361
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod, G.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes, Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch.
A:Reference number: A69250; MUID:98049343
A:Accession: D69361
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

Db 126 AAKANVVKCFSSVGS-GLAVNPOLKGEKPSMFICGNDSSKQKRIKILDTFGMDTEDMK 184
QY 195 LSSAREIENPLRLFTLW 212
Db 185 VEARAIEPLCI---LW 198

RESULT 12

Conserved hypothetical protein XF1737 [Imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: G82642

R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequences 406, 151-157, 2000
X:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: AB2515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: G82642

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-242 <SIM>
A:Cross-references: GB:AE003997; GB:AE003849; NID:99106805; PIDN:AAF4546.1; GSPDB:GN001

A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A. Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H. de-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
Submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kempfer, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig Chad, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, H. A. Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracosta, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A. Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Saneili, R.V.; Sawasak A. Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1737

Query Match 4.6%; Score 116.5; DB 2; Length 242;

Best Local Similarity 24.0%; Pred. No. 0.089;

Matches 46; Conservative 40; Mismatches 85; Indels 21; Gaps 7;

QY 28 ARKVTGIGSGDFAKSLTIRLCGYHVVIGSRNP---KFASEFPFHYVDVTHHEDAL 83

Db 30 AAPMRIGVIGAGSLGTGVLWAKGHEVWFSRNPDKLEAMARELEPR-ASVGQPLAAT 88

QY 84 TKTNIIFVAIHREHYTSLM-DLRHLVYKILIDVSNMNMINO---YPSRN---AEYLAS 135

Db 89 EFGTVLLAVPEALPQVGRDLSAYRGKIVLDTNPMGASSADVYRREARELGAQVYAK 148

QY 136 LPDPSLIVGFNVASAMALQGPDKASRO---VYICSNNTIOARQVTEIARQLFTPI 190

Db 149 YMPGALVIAFSAVDATVY---TSASRGRGRIQMLASDDAEAMKVAEGLVADACDPV 205

QY 191 DLGSLSSAREIE 202

Db 206 IVCNLAAMASFQ 217

RESULT 13

AB3182
Conserved hypothetical protein Atus183 [Imported] - Agrobacterium tumefaciens (strain C5
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AB3182

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I. erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell J.; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kan ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AB3182

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-198 <KUR>

A:Cross-references: GB:AE008687; PIDN:AA45872.1; PID:q17743615; GSPDB:GN00188

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atus183

A:Genome: plasmid

Query Match 4.6%; Score 116; DB 2; Length 198;

Best Local Similarity 22.7%; Pred. No. 0.076;

Matches 50; Conservative 41; Mismatches 81; Indels 48; Gaps 10;

QY 31 VYGVIGSGDFAKSLTIRLCGYHYI-GSRPKRASEFFPV---VDYTHHEDALTKT 86

Db 1 MVTGIIAGAGIGAFATGACIGIAVIANSRGPESTLALVSKLSTIRAGSVPEAAQA 60

QY 87 NIIFVAIHREHYSLMD---LRHLVG-KILIDVSNMNRINOY-----PESNAEYLA 134

Db 61 -YLVAVP-----WSKIPGALAGLFGDRIYIDANNSTIEAPLYRPAIDGRTSDIFF 112

QY 135 LPDPSLIVGFNVASAMALQGPDKAS-----ROYICSNNTIOARQVTEIARQL 185

Db 113 ALVPGARVYAFN-----HLTPKOLSGDPHSEGRVLYFGSDMDRAKAEVGAIDRI 165

QY 186 NFPIIDGSLSSAREIENPLRLFTLRGPVVAISLAF 225

Db 166 GFGIDLGGLPGVSGQLQFP-----GSPFALNLVKT 197

RESULT 14

JC5653
G protein-coupled receptor - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 17-Nov-1997 #sequence_revision 17-Nov-1997 #text_change 21-Jul-2000
C:Accession: JC5653
R:Ferrer-Martinez, A.; Felipe, A.; Mata, J.F.; Casado, F.J.; Pastor-Anglada, M. Biochem. Biophys. Res. Commun. 238, 107-112, 1997
A:Title: Molecular Cloning of a bovine renal G-protein coupled receptor gene (bngR):
A:Reference number: JC5653; MUID:97445134
A:Accession: JC5653
A:Molecule type: mRNA
A:Residues: 1-361 <PER>

A:Cross-references: GB:U88366; NID:92827875; PIDN:AAC05611.1; PID:92827876
A:Experimental source: renal epithelial cell
C:Superfamily: G protein-coupled receptor 4
C:Keywords: glycoprotein; phosphoprotein
F:24-46/Domain: transmembrane #status predicted <TM1>
F:58-79/Domain: transmembrane #status predicted <TM2>
F:96-117/Domain: transmembrane #status predicted <TM3>
F:137-159/Domain: transmembrane #status predicted <TM4>
F:189-207/Domain: transmembrane #status predicted <TM5>
F:229-253/Domain: transmembrane #status predicted <TM6>
F:265-289/Domain: transmembrane #status predicted <TM7>
F:318-62/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:134,218,293,313/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status
F:204,221,328/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status

Query Match 4.5%; Score 114.5; DB 2; Length 361;

Best Local Similarity 20.4%; Pred. No. 0.21;

Matches 56; Conservative 51; Mismatches 86; Indels 81; Gaps 13;

QY 255 VNKTL-PIVATITLISVYAGLIAAAYQLYGTRRFPPEWETWQCKRQGLSFFFA 313

Db 19 IHQTLAPVYV---WLVAVGFANCLSLYYG-----YLD--IKARELGVYLCNLT 64

```

OY 314 MHHVAVSICLPMRRSERLFLNMAVQOVHANIENSNEE-----EVMHIEHYISFGIMSG 369
Db 65 VADFLYICSLP-----FWLQYLVQHDH-----WSDHDSOCVCICILYENIYI-SVG 110
OY 370 LLSLLAVTISPSVSNALNMRRESFISQSTLGYAL-----LISFPHVLIV-----GKK 416
Db 111 FLCCSIDIRYLAVAHPRFHQFRLTKAMGSVALIWKELTISYFLMBEVEDADRHR 170
OY 417 RAFEI-----EYRRFTPPNPFVALVLSI-----VI 443
Db 171 VCFEHYPLEPRQGINVYRFLVGLFPIICLLLASRGILRAVRRSHOTOKSRKDOIORLV 230
OY 444 LGGIILFLPCL-SRKLRKIKKGKESQPLDEGI 475
Db 231 LSTVIVIFLACFLPYHLLVLLVNSLMESSCDRAKI 264

```

Gencore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2002, 02:12:50 ; Search time 37.74 Seconds
(without alignments)
502.718 Million cell updates/sec

Title: US-09-802-520-1

Perfect score: 2534
Sequence: 1 MESISMWGSKSLSETCLPN.....LEEGIGTIPHSPEPTYM 490

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	788	31.1	339	1 STEA_HUMAN	Q9UHE8 homo sapien
2	133	5.2	223	1 FARE_METJA	O58896 methanococc
3	121.5	4.8	629	1 FARE_YEAST	Q12333 saccharomyc
4	120.5	4.8	232	1 FARE_METTH	O26350 methanobact
5	118.5	4.7	365	1 GPE8_HUMAN	Q15743 homo sapien
6	110	4.3	695	1 FSHR_MACFA	P32212 macaca fasc
7	108	4.3	712	1 FPRE6_YEAST	Q12473 saccharomyc
8	107	4.2	1769	1 YUK9_YEAST	P42945 saccharomyc
9	106	4.2	695	1 FSHR_HUMAN	P23945 homo sapien
10	104	4.1	396	1 SOTRB_ECO57	P58529 escherichia
11	104	4.1	396	1 SOTRB_ECO1	P31122 escherichia
12	103.5	4.1	268	1 YD49_AQUAE	O67364 aquifex aeo
13	103.5	4.1	465	1 NPPI_MOUSE	O61983 mus musculu
14	103.5	4.1	521	1 YR25_CAEEL	Q10934 caenorhabdi
15	103	4.1	695	1 FSHR_PIG	P45059 sus scrofa
16	102	4.0	693	1 FSHR_CHICK	P79763 gallus gali
17	101	4.0	393	1 NOCC_ARATH	P56755 arabidopsis
18	101	4.0	393	1 NOCC_SPIOL	O9m315 spinacia ol
19	100	3.9	476	1 RF3_SACBA	P05512 saccharomyc
20	99	3.9	458	1 SSUL_YEAST	P41930 saccharomyc
21	98.5	3.9	452	1 NUD4M_BRALA	O79421 branchiosto
22	98.5	3.9	3411	1 POLG_YEYV1	P03314 y genome po
23	98.5	3.9	3411	1 POLG_YEYV2	P19901 y genome po
24	96.5	3.8	686	1 FREL_YEAST	P32791 saccharomyc
25	96	3.8	261	1 PHSC_ECOLI	P77409 escherichia
26	96	3.8	447	1 NUD4M_APILI	P34853 apis mellif
27	96	3.8	695	1 FSHR_BOVIN	P35376 bos taurus
28	95.5	3.8	393	1 NOCC_RHISA	P12132 oryza sativ
29	94.5	3.7	433	1 NUD4M_RHISA	O99825 rhinocerot
30	94.5	3.7	452	1 NUD4M_BRAL	O47423 branchiosto
31	94.5	3.7	492	1 SECY_CYAPA	P25014 cyanophora
32	94.5	3.7	1228	1 ATC_PLAFK	O08853 plasmodium
33	94	3.7	891	1 VP4A_FOWPV	O93559 fowlpox vir

ALIGNMENTS

RESULT ID	STEA_HUMAN	STANDARD:	PRT:	339 AA.	
34	93.5	3.7	499	1	GSHR_PLAF7
35	93.5	3.7	1233	1	VLI_REOVD
36	93	3.7	388	1	HMC3_DESVH
37	93	3.7	564	1	HXTC_YEAST
38	93	3.7	692	1	FSHR_RAT
39	93	3.7	695	1	FSHR_SHEEP
40	92.5	3.7	379	1	CYB_MIRLE
41	92.5	3.7	416	1	YH04_AQUAE
42	92.5	3.7	477	1	BAIG_EUBSP
43	92.5	3.7	490	1	GTRI_CHICK
44	92.5	3.7	1120	1	KEPA_ECOLI
45	92.5	3.7	2037	1	FASI_CANAL

RESULT 1
STEA_HUMAN
AC Q9UHE8: Q95034; STANDARD: PRT: 339 AA.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Six transmembrane epithelial antigen of prostate.
GN STEAP OR STEAP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:20056277; PubMed:10588738;
RA Hubert R.S., Vivanco I., Chen E., Rastegar S., Leong K.,
RA Mitchell S.C., Madraswala R., Zhou Y., Kuo J., Raitano A.B.,
RA Jakobovits A., Saffran D.C., Afar D.E.H.;
RT "STEAP: a prostate-specific cell-surface antigen highly expressed in
human prostate tumors";
RT Proc. Natl. Acad. Sci. U.S.A. 96:14523-14528(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Abu-Threideh J., Stoneking T., Langston Y., Maupin R.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PROSTATE TUMORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AF186249; AAF17479.1; -
CC EMBL: AC005053; AAC79150.1; ALT_INT.
CC EMBL: AC004969; AAD15620.1; ALT_INT.
CC MIM: 604415; -
CC Transmembrane; Antigen.
CC
CC TRANSMEM 71 91 POTENTIAL.
CC TRANSMEM 119 139 POTENTIAL.
CC TRANSMEM 164 184 POTENTIAL.
CC TRANSMEM 218 238 POTENTIAL.
CC TRANSMEM 258 278 POTENTIAL.
CC TRANSMEM 291 311 POTENTIAL.
CC SEQUENCE 339 AA; 39851 MW; 55443A170C870387 CRC64;
Query Match 31.1%; Score 788; DB 1; Length 339;
Best Local Similarity 54.8%; Pred. No. 4,9e-48;
Matches 142; Conservative 53; Mismatches 64; Indels 0; Gaps 0;

OY 208 LFTLMGPPVVAISLAEFFLVSFVADVIHPYARNQSDPFYKIPILVKTLPVAVITLL 267

DR PRINTS: PR00237; GPCR_RHO�PSN.
 DR PROSITE; PS00237; G-PROTEIN_RECEP_F1_1; 1.
 DR PROSITE; PS50262; G-PROTEIN_RECEP_F1_2; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein.
 FT DOMAIN 1 21
 FT TRANSSEM 22 46
 FT TRANSSEM 47 58
 FT TRANSSEM 59 80
 FT TRANSSEM 81 95
 FT TRANSSEM 96 117
 FT TRANSSEM 118 136
 FT TRANSSEM 137 158
 FT TRANSSEM 159 183
 FT TRANSSEM 184 203
 FT TRANSSEM 206 228
 FT TRANSSEM 229 249
 FT TRANSSEM 250 263
 FT TRANSSEM 264 284
 FT TRANSSEM 285 365
 FT DISULFID 94 172
 FT CARBOHYD 3 3
 FT CARBOHYD 8 8
 FT CONFLICT 140 140
 FT CONFLICT 142 142
 FT CONFLICT 142 142
 SQ SEQUENCE 365 AA; 41076 MM; 05919AFD5B842CCD CRC64;

Query Match 4.7%; Score 118.5; DB 1; Length 365;
 Best Local Similarity 21.0%; Pred. No. 0.24;
 Matches 57; Conservative 46; Mismatches 92; Indels 77; Gaps 11;

QY 255 VNKL-PIVATLTLSTLYAGLAAATQLYGTYRRPPLEWLOCRKGLISFFFA 313
 DB 19 IHHQLAVVYTVL-----VVGFPANCLSLYFG-----YLQ--IKARNELGYLNLMT 64
 QY 314 MVHAYSLCLPMRSEKRYLFLNMAVQOVHANIENSWNEEVEWREMTSEFGINSGLISL 373
 DB 65 VADLFYICSLPFW-----LDYVLQHDWMSHGDISCOVGLILENYIT-SVGFLLCC 114
 QY 374 LAVTSPVSNAALNMRSEFSTLGYVA-----LLISTP----- 408
 DB 115 ISVDRYLAHNPFRFHQFRIKAAGVSVYIAWAKELLSTYFLMHEVIEDENGRVCFE 174
 QY 409 HVLVYGMKRAFEERYRPPNFVALVLPST-----VILG 445
 DB 175 HYPQAOQRAI--NYRFLVGLFPLICLLASVGIILAVRSHGTOKSRKQIORIVLS 232
 QY 446 KIILFPCI--SRKLRIKKGMESQFLBEGI 475
 DB 233 TVVIFLACFLPYHVLVLRVSYEASCDPAKGV 264

RESULT 6
 FSHR_MACFA STANDARD; PRT: 695 AA.
 ID F32212;
 DC 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Follicle stimulating hormone receptor precursor (FSH-R) (Follicitropin receptor).
 GN FSHR.
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Cercopitheciidae; OC Cercopitheciidae; Macaca.
 NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=94071854; PubMed=7504463;
 RA Gromoll J., Dankbar B., Sharma R.S., Nieschlag E.;
 "Molecular cloning of the testicular follicle stimulating hormone

RT receptor of the non human primate Macaca fascicularis and
 RT Identification of multiple transcripts in the testis.";
 RL Biochem. Biophys. Res. Commun. 196:1066-1072(1993).
 CC -1- FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYLATE CYCLASE.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC FSH/LSH/TSR SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 6 LEUCINE-RICH REPEATS (LRR).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; X74454; CA52463.1; .
 CC PIR; S36452; S36452.
 CC PIR; JN0898; JN0898.
 CC HSSP; P23945; 1XUN.
 CC GCRDB; GCR_0653; .
 DR InterPro; IPR000276; GPCR_Rhodopsn.
 DR InterPro; IPR001611; LRR.
 DR InterPro; IPR000372; LRR_Nterm.
 DR Pfam; PF00001; 7tm_1; 1.
 DR Pfam; PF00560; LRR; 4.
 DR Pfam; PF01462; LRRNT; 1.
 DR PRINTS; PR00373; GLXCHROMNER.
 DR PRINTS; PR01143; FSHRECEPTOR.
 DR SMART; SM00013; LRRNT; 1.
 DR PROSITE; PS00237; G-PROTEIN_RECEP_F1_1; 1.
 DR PROSITE; PS50262; G-PROTEIN_RECEP_F1_2; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein; signal;
 KW phosphorylation; Repeat; Leucine-rich repeat.
 FT SIGNAL 1 17
 FT CHAIN 18 695
 FT DOMAIN 18 366
 FT TRANSSEM 367 387
 FT DOMAIN 388 398
 FT TRANSSEM 399 421
 FT DOMAIN 422 443
 FT TRANSSEM 444 465
 FT DOMAIN 466 485
 FT TRANSSEM 486 508
 FT DOMAIN 509 528
 FT TRANSSEM 529 550
 FT DOMAIN 551 573
 FT TRANSSEM 574 597
 FT DOMAIN 598 608
 FT TRANSSEM 609 630
 FT DOMAIN 631 695
 FT REPEAT 44 68
 FT REPEAT 69 93
 FT REPEAT 119 143
 FT REPEAT 170 192
 FT REPEAT 193 216
 FT REPEAT 218 240
 FT DISULFID 442 517
 FT CARBOHYD 191 191
 FT CARBOHYD 199 199
 FT CARBOHYD 293 293
 FT CARBOHYD 318 318
 SQ SEQUENCE 695 AA; 78343 MM; 0D60A233729B5250 CRC64;

Query Match 4.3%; Score 110; DB 1; Length 695;
 Best Local Similarity 16.5%; Pred. No. 2;
 Matches 84; Conservative 79; Mismatches 151; Indels 196; Gaps 17;

QY 79 HEDALTNITLFAVIAHREHYTSLMDLRHLVG-----KILIDVSNMNR 121


```

Db 98 HEIRIKANNL-LYINPEAFONLPNRLYLISNTGIKHLPDVAKHSFOKVLIDIDNIN 156
OY 122 INQPESSNAEYLASFPPDSLYKGFENVVSAMALQGPKASROYICSNINICARQOYIEL 181
Db 157 IH-----TIERNSEFVGLSFESVILMT-----NKGIOEIHNC----- 189
OY 182 AROLNFIPIIDGLSSAREIENPLRLFTLMRGPPVVAISLAFFLYSPVRDVIHPYAR 241
Db 190 ----FNCTQDELNLSDNNMLLEELPNDVFHGASGPVILDISIRIRHSLPSTIGLENKLLRA 246
OY 242 NOOSDFYKPIPIEIVNKTPIVAITLLSLVY----- 271
Db 247 RSTYNNKRLP-----SLEKVALMEASLTFYPHSCAFAMWRQISLHPICNKSILROEV 301
OY 272 -----LAGLLAAAYOLITYGTRRF-----PPMLETWLOCKROIG 306
Db 302 DYMOTQRGORSSLAEDNESSYSKGFDMTAEFDYDLCEVVDVTCSPKPDAPNCPEDIIIG 361
OY 307 -----LSFFPAMVHYA-----YSICLPMRSEERYFLNMAVYQOVHANIEHS 348
Db 362 YNLRVLIWFISLITATGNIIVLVLTTSQYKLYVP-----RFLMCNLAFAD----- 408
OY 349 WNEEVEWRIEMYISGIMSLGLSLAVTSIPSVSNALNMR-----FSFIQSTIG 399
Db 409 -----LCIGIYLLLASVDIHTKSOYHNYAIDMQGAGCDAGFTVFASLS 456
OY 400 -----YVALLISTFH-----VLITGKRAREEYRYTPPNEY-LAL 436
Db 457 VYTLAITLERWHTTHAMQLDCKVHRHAASVMWGWTFAPFAALFPIFGISSYMKVSI 516
OY 437 VLP-----SIVILGKIILFLPC 453
Db 517 CLPMIDSPLSQLYWMSLVLANVLAERVIC 546

```

```

RESULT 7
FREQ_YEAST STANDARD: PRT: 712 AA.
AC 012473:
Db 15-DEC-1998 (Rel. 37, Created)
Dt 15-DEC-1998 (Rel. 37, Last sequence update)
Dt 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ferric reductase transmembrane component 6 precursor (EC 1.6.99.13)
GN FERRIC-CHLATEL REDUCTASE 6.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Saccharomycetes; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OX Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
RN NCBI_TaxID=4932;
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RA Wedler H., Wandubt R.;
RL Submitted (JAN-1995) to the EMBL/Genbank/DBJ databases.
CC -1 CATALYTIC ACTIVITY: NADH + 2 Fe(3+) = NAD(+) + 2 Fe(2+).
CC -1 COFACTOR: FAD (PROBABLE).
CC -1 SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1 SIMILARITY: BELONGS TO THE FRE / CYBB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 247973; CAA88006.1; -
DR EMBL: 273156; CAA97503.1; -
DR SGD: S0003974; FRE6.
DR InterPro: IPR002916; Ferric_reduct.
DR Pfam: PF01794; Ferric_reduct; 1.

```

```

KW Oxidoreductase; Electron transport; Transmembrane; Iron transport;
KW FAD; NAD; Glycoprotein; Signal; Multigene family.
FT SIGNAL 1 17
FT CHAIN 18 712
FT NP_BIND 493 499 FAD (POTENTIAL).
FT TRANSMEM 168 188 POTENTIAL.
FT TRANSMEM 245 265 POTENTIAL.
FT TRANSMEM 288 308 POTENTIAL.
FT TRANSMEM 329 349 POTENTIAL.
FT TRANSMEM 361 381 POTENTIAL.
FT TRANSMEM 388 408 POTENTIAL.
FT TRANSMEM 417 437 POTENTIAL.
FT TRANSMEM 478 498 POTENTIAL.
FT TRANSMEM 550 570 POTENTIAL.
FT DOMAIN 553 556 POLY-LEU.
FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 450 450 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 655 655 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 712 AA; 81989 MW; 5224F12B51544BAA CRC64;

```

Query Match 4.3%; Score 108; DB 1; Length 712;
 Best Local Similarity 18.5%; Pred. No. 2.8;
 Matches 74; Conservative 52; Mismatches 114; Indels 160; Gaps 17;

```

OY 88 IIFVAIHREHYTSL-----MDRLHLVGKILIDVSNMGRINOYPESNAEYLA 136
Db 183 IIAVFPMHSHYNGLNALFASREFVNYIRGHFVLPTFLVD-----KHANHEFLNVEFTGL 238
OY 137 FPPDSLYKGFENVVSAMALQGPKASROYICSNINICARQOYIELAROLNFIPIIDGLSS 196
Db 239 MPNSL-----EAW----- 246
OY 197 SAREIENPLRLFTLMRGPPVVAISLATEFEFL-YSEFVDVIHPYARNNOOSDFYKPIEIV 255
Db 247 -----IIEGYTLANIIFLSISYIIDPYNLIFNLSHQFRL---LA 284
OY 256 NKTPIVAITLLSLVYLAGLLAAAYQLYGTYKRRPFWLETWLOCRKOLGLISFFPAMY 315
Db 285 DRS-GILAFQFPLIILIFARNSEFLFLGVKRNSE-----ISPHKMGIRIMVLAFTI 336
OY 316 H-VAYSLCLPMRSEERYFLNMAVYQOVHANIEHSNDEEVEWRIEMYISGIMSLGLSL 374
Db 337 HSLSYSL-----FAIINNAFK-----ISNK-----QLYWKFGIASITVLCVL 373
OY 375 AVTSIPSVSNALNMRSEFSFIQSTIGYVALLI---STFHVLIY-GMKRAPEEYRYRFTTP 430
Db 374 LVLSLGIYVR-----RHFEFLYTHIILALLFLFCCHQKHVFNKME----- 416
OY 431 NEVLALVPSIVILGKIIFLPCISIRKLRIKKMKESQF 470
Db 417 ----WIWVSLIHW-----LEKRLFRIMWLDG 439

```

```

RESULT 8
YK9_YEAST STANDARD: PRT: 1769 AA.
AC P42945;
Dt 01-NOV-1995 (Rel. 32, Created)
Dt 01-NOV-1995 (Rel. 32, Last sequence update)
Dt 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 200.0 kDa protein in G2F3-IME2 intergenic region.
GN YJL109C OR J0808.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Saccharomycetes; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OX Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
RN NCBI_TaxID=4932;
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;

```

RX MEDLINE-96090136; PubMed-7483851;
 RA Rasmussen S.W.;
 RT "A 37.5 kb region of yeast chromosome X includes the SWE1, MEP2, GSH1
 RT and GSD3 genes, a TCP-1-related gene, an open reading frame similar
 RT to the DAL80 gene, and a tRNA(Arg).";
 RL Yeast 11:873-883(1995).
 CC - SIMILARITY: BELONGS TO THE BAP28 FAMILY.
 CC - SIMILARITY: CONTAINS 1 HEAT REPEAT.
 CC -
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X85021; CA59385.1; -;
 DR EMBL: Z49384; CA89404.1; -;
 DR SGD: S0003645; YJL109C.
 DR InterPro: IPR000357; HEAT_repeat.
 DR PROSITE: PS50077; HEAT_REPEAT; 1.
 KW Hypothetical protein.
 FT REPEAT 1729 1767 HEAT.
 SQ SEQUENCE 1769 AA; 20080 MW; 064480DD1249B241 CRC64;

Query Match 4.28; Score 107; DB 1; Length 1769;
 Best Local Similarity 19.98; Pred. No. 9.4; Indels 144; Gaps 22;

Matches 106; Conservative 79; Mismatches 205; Indels 144; Gaps 22;

QY 17 CLPFGINGIDARKVT-VGVIGSGDFAKSLTIRLIRCGYHVIGSRNPKAFSEFPFPHVD 75
 DB 163 CLSNFVSEKPPALTMKLFNDMFLKLYTSYDQCIKH-----NATYNO----- 209
 QY 76 VTHHEDALFTNITFEVAIHREHTSLMDLRLHVLGKILIDVSNMKNRNP-----ESNAE 131
 DB 210 -----LFTTCCFIV-----VAFNSNDEKLNQVLPILLETSK 244
 QY 132 YLASFPD-SLIVGFVNSAMALQPKDASROYIICSNNOQAOVIELAROLNTP 190
 DB 245 LKASKSDQCIANTLIVFATNLPKKTITIAMETILSLDQKE-----AKSALLTI 299
 QY 191 D--LGSLSAREIENPLRLFTMRGPVVAISLATPF-----FLYSEYVDVIHP 238
 DB 300 CKLFQFLKGGQNDQLSKIFLFDSDK-PTVYSILTFLEDKEDKPYCDKFTISYRSIAR- 357
 QY 239 YARNO-----QSDFKPIEIVNKTPLVATITLSLVLAGLAAAYOLYGTKRPRPPW 294
 DB 358 YDNRKLTITLSLKKIKLEERE-----VRLITDLYLSELEDSOL-----VELEYE 407
 QY 295 L---ETWLCROKOLGLSEFFAM-----VHVAYSICLPMRSEYRLFNMA 337
 DB 408 ISINEDLVKCLKSLGTLGELFEIRLTSLPTNADVNTDIYKQSDPVEVTKKOTASFO 467
 QY 338 YQOVHANIEENSWN---EEFVWRIENTISFGINSLG-----LISL 374
 DB 468 FLPKHSELITVTWVSMLETFGERKRYLSLFTETAIKGYKASSPLTSEFTTLESRIFFL 527
 QY 375 AVTSIPSVSNAUMMRSEFQSITGLYVALLISTFHVILYGMKRAFEERYRFTIPPEVL 434
 DB 528 RVTISPAAPALK-----LISLNNAKY--INSIEKEVINFTLVPLIC 569
 QY 435 ALVLPISVIGKILILPCLISRLKRIKGMKESQFLEEGIG---TIPVSP 485
 DB 570 ALMDASIKVTVGKILISLA-----KRSTKHVFLSDKLYGENTYIPMLNR 617

RESULT 9
 FSHR_HUMAN
 ID FSHR_HUMAN STANDARD; PRT; 695 AA.
 AC P23945;
 DT 01-MAR-1992 (Rel. 21, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Follicle stimulating hormone receptor precursor (FSH-R) (Fol11tropin
 DE receptor).
 GN FSHR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 NCBI_TaxID=9606;
 OX [1]
 RP SEQUENCE FROM N.A.
 RP TISSUE-Ovary.
 RC MEDLINE-91222171; PubMed-1709010;
 RA Muegish T., Nakamura K., Takakura Y., Itoh Y., Igarashi M.;
 RT Cloning and sequencing of human FSH receptor cDNA.
 RL Biochem. Biophys. Res. Commun. 175:1125-1130(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RP TISSUE-Testis;
 RC MEDLINE-93246012; PubMed-1301382;
 RA Kelson C.A., Cheng S.V., Nugent N.P., Schweichardt R.L.,
 RA Rosenthal J.L., Overton S.A., Wands G.D., Kuzaja J.B., Luchette C.A.,
 RA Chappel S.C.;
 RT The cloning of the human follicle stimulating hormone receptor and
 RT its expression in COS-7, CHO, and Y-1 cells.
 RL Mol. Cell. Endocrinol. 89:141-151(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Tilly L.T., Alhara T., Nishimori K., Jai X.-C., Billig H.,
 RA Kowalski K.I., Perlas E.A., Hsueh A.J.;
 RL Submitted (xxx-1992) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 1-342 FROM N.A.
 RP TISSUE-Testis;
 RC MEDLINE-93075197; PubMed-1359889;
 RA Gromoll J., Gudermann T., Nieschlag E.;
 RT Molecular cloning of a truncated isoform of the human follicle
 RT stimulating hormone receptor.
 RL Biochem. Biophys. Res. Commun. 188:1077-1083(1992).
 RN [5]
 RP SEQUENCE OF 1-51 FROM N.A.
 RC MEDLINE-95011044; PubMed-7926278;
 RA Gromoll J., Dankbar B., Gudermann T.;
 RT Characterization of the 5' flanking region of the human follicle-
 RT stimulating hormone receptor gene.
 RL Mol. Cell. Endocrinol. 102:93-102(1994).
 RN [6]
 RP 3D-STRUCTURE MODELING OF 49-228
 RC MEDLINE-96363672; PubMed-8747461;
 RA Jiang X., Dreano M., Buckler D.R., Cheng S., Ythier A., Wu H.,
 RA Hendrickson W.A., el Tayar N.;
 RT Structural predictions for the ligand-binding region of glycoprotein
 RT hormone receptors and the nature of hormone-receptor interactions.
 RL Structure 3:1341-1353(1995)
 CC - FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY
 CC OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
 CC ADENYLATE CYCLASE.
 CC - SUBCELLULAR LOCATION: Integral membrane protein.
 CC - ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM. ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC - TISSUE SPECIFICITY: SERTOLEI CELLS AND OVARIAN GRANULOSA CELLS.
 CC - SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC FSH/LSH/TSH SUBFAMILY.
 CC - SIMILARITY: CONTAINS 5 LEUCINE-RICH REPEATS (LRR).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

```
DR EMBL: M65085; AAA52477.1; -
DR EMBL: S59900; AAB26480.1; -
DR EMBL: M95489; AAA52478.1; -
DR EMBL: X68044; CAA48179.1; -
DR EMBL: S73199; AAB32071.1; -
DR PIR: JN0122; JN0122.
DR PDB: 1XUN; 15-MAY-97.
DR GCRDB: GCR_0071; -
DR GCRDB: GCR_0404; -
DR GCRDB: GCR_0588; -
DR GCRDB: GCR_0690; -
DR MIM: 136435; -
DR InterPro: IPR000276; GPCR_Rhodopsn.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR00372; LRR_Nterm.
DR Pfam: PF00001; 7tm_1; 1.
DR Pfam: PF00560; LRR; 4.
DR Pfam: PF01462; LRRNT; 1.
DR PRINTS: PR00373; GLYCHOMONER.
DR PRINTS: PR01143; FSHRECEPTOR.
DR SMART: SM00013; LRRNT; 1.
DR PROSITE: PS00237; G_PROTEIN_RECP_F1_1; 1.
DR PROSITE: PS50262; G_PROTEIN_RECP_F1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Phosphorylation; Repeat; Leucine-rich repeat; Alternative splicing;
KW 3D-structure.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 695 FOLICLE STIMULATING HORMONE RECEPTOR.
FT DOMAIN 18 366 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 367 387 1 (POTENTIAL).
FT DOMAIN 388 398 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 399 421 2 (POTENTIAL).
FT DOMAIN 422 443 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 444 465 3 (POTENTIAL).
FT DOMAIN 466 485 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 486 508 4 (POTENTIAL).
FT DOMAIN 509 528 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 529 550 5 (POTENTIAL).
FT DOMAIN 551 573 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 574 597 6 (POTENTIAL).
FT DOMAIN 598 608 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 609 630 7 (POTENTIAL).
FT DOMAIN 631 695 CYTOPLASMIC (POTENTIAL).
FT REPEAT 69 93 LRR 1.
FT REPEAT 119 143 LRR 2.
FT REPEAT 170 192 LRR 3.
FT REPEAT 193 216 LRR 4.
FT REPEAT 218 240 LRR 5.
FT DISULFID 442 517 BY SIMILARITY.
FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPPLIC 224 285 MISSING (IN SHORT ISOFORM).
FT VARSPPLIC 342 695 MISSING (IN SHORT ISOFORM).
FT CONFLICT 13 13 S -> R (IN REF. 4).
FT CONFLICT 112 112 N -> T (IN REF. 1).
FT CONFLICT 197 198 EL -> AV (IN REF. 1).
FT CONFLICT 295 295 S -> P (IN REF. 4).
FT CONFLICT 307 307 T -> A (IN REF. 1).
FT CONFLICT 680 680 N -> S (IN REF. 1).
SO -SEQUENCE 695 AA; 78294 MM; 723B8EF1F76D2CD5 CRC64;
```

Query Match 4.2%; Score 106; DB 1; Length 695;
Best Local Similarity 16.4%; Pred. No. 3.8;
Matches 84; Conservative 81; Mismatches 148; Indels 198; Gaps 18;

```
OY 79 HEDALTKNIIIVAIHREHYTSLSMDRLHLV-----KIILDSNNMR 121
DB 98 HIRIEKANNL-LYINPEFOMLPNIQYLISNTGIKHLPDVHKIHSLOKVLIDODINT 156
OY 122 INQYPSNAEVLASLEPDSLIYKGFNVASAMALQGPKDSROYICSNINQARQVIEL 181
```

```
DB 157 IH-----TIERNSPVGLSFESVILML-----NKNGIOEIHNCA----- 189
OY 182 AROLNFIPIIDGLSSARELENIPLRLFTLMKGPVVAISLATFFLYSFVRVYHPYAR 241
DB 190 ---FNGCTOLDELNLSDNNNIEELPNDVFGASGPVILDISRIRHSLPSYGLENLKRLA 246
OY 242 NQOSDFYKPIELIVNKT-LPIVATITLSLY----- 271
DB 247 RSTYNLKLKP-----TLKVALMEASLTPYSHCCAFANWRQISELPHICKSLROE 300
OY 272 -----LAGLLAAAYOLYGYTKYRRF-----PPMLETLQCRKOL 305
DB 301 VYIMQTRGQRSLAEDNSSYSGPDMYTTERDYDLCEVVDYWCSPDPANPCEDIM 360
OY 306 G-----LSFFPRAMHVA-----YSLCLPMRSEYFLFNAYQOYHANEN 347
DB 361 GYNILRLVLMFISLITAGNIIVLVITTSQYKLTVP-----RFLMCNLAFAAD----- 408
OY 348 SNNEEVMRLIEMYSISGMSLGLSLAATVSIPIVSNALMWR-----PSFIQSTL 398
DB 409 -----LCIGITVLLIASVDIHTKSOYHNATIDMGTGACDAAGFTVASEL 455
OY 399 G---YVALLISFTH-----VLTYGKRAREEYRYRPPNFV-LA 435
DB 456 SVYTLVAITLERHWTTHAMQLDCKYQLRHAASVVMGMTEFAAALPFIIGISSYMKVS 515
OY 436 LVLP-----STVILGKITLFLPC 453
DB 516 ICLPMDIDSPLSQLYVMSLVLNVLAFVVIC 546
RESULT 10
SOTB_ECO57 STANDARD; PRT; 396 AA.
ID SOTB_ECO57
AC P58529;
DT 01-MAR-2002 (rel. 41, Created)
DT 01-MAR-2002 (rel. 41, Last sequence update)
DT 01-MAR-2002 (rel. 41, Last annotation update)
DE Sugar efflux transporter.
GN SOTB OR Z2173 OR ECS2135.
OS Escherichia coli O157:H7.
OC Escherichia; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postfal G., Hackett J., Link S., Boutin A., Shao Y., Miller L.,
RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blatter F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohnishi E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Tida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -I- FUNCTION: Involved in the efflux of sugars. The physiological role
CC may be the reduction of the intracellular concentration of toxic
CC sugars or sugar metabolites. Transports L-arabinose and to a
CC lesser extent IPTG. Seems to contribute to the control of the
CC arabinose regulon (by similarity).
```

CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (Probable).
 CC -1- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN AS
 CC THE DRUG RESISTANCE TRANSLOCASE FAMILY). SOTB (TC 2.A.1.2)
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: AEO05352; AAG56236.1; -
 CC KX EMBL: AF002557; BAB35558.1; -
 CC KM Transport: Sugar transport; Transmembrane; Inner membrane.
 CC FT TRANSMEM 15 35
 CC FT TRANSMEM 50 70 POTENTIAL.
 CC FT TRANSMEM 81 101 POTENTIAL.
 CC FT TRANSMEM 103 123 POTENTIAL.
 CC FT TRANSMEM 136 156 POTENTIAL.
 CC FT TRANSMEM 170 190 POTENTIAL.
 CC FT TRANSMEM 209 229 POTENTIAL.
 CC FT TRANSMEM 246 266 POTENTIAL.
 CC FT TRANSMEM 275 295 POTENTIAL.
 CC FT TRANSMEM 319 339 POTENTIAL.
 CC FT TRANSMEM 333 353 POTENTIAL.
 CC FT TRANSMEM 364 384 POTENTIAL.
 CC SQ SEQUENCE 396 AA; 42568 MW; CB6A209B0B6DAF0 CRC64;

Query Match 4.1%; Score 104; DB 1; Length 396;
 Best Local Similarity 18.8%; Pred. No. 2.7;
 Matches 85; Conservative 72; Mismatches 148; Indels 146; Gaps 20;

QY 31 VTGVVIGSGPFAKSLTRLRGCHVAVIGSRNPKFSEFPVVDVTHHEDALTKTNIIF 90
 DB 31 VPGVLT--SDIAQSFHMQTQVQIMLTIVAVVVALMSLPFMLTMSOVERRKLCLPEVVF 88
 QY 91 VAHREHYTSIMDLRHLVGLIIDVSNMNRINQYPSNMEYLASLPDLSLIVGPNVVS 150
 DB 89 IASIVLSFLS-WSTFVTVLISRI-----GVAFAHAIF-----WSITRA 123
 QY 151 AMAALQCP--KDSROYVYCSNNIQAQOVELAR-----OLNFIPIIDGLSL----- 195
 DB 124 SLATRMAPAGRAQALSLIATGATLAWLGLPLGRIVGQYRGWMTFPAIGALITLTL 183
 QY 196 -----SSAREINLPLRLFTLMRGVVAISLAT-----FFLVISFVADVHPY 239
 DB 184 LKILPLLPSEHSGSLSLP---LFRPALMSIYLLTVVVVTAHTTAVSY---IEPF 235
 QY 240 ARN---QOSDFYKIPRIEIVNKTLPVATLTLSTLYLAGLLAAAYQLYYGFYRRFPWLE 296
 DB 236 VQNTAGSANP-----ATALLLLLGAGCITIGSVFGLKGNAYAS----- 274
 QY 297 TWIQCROQLGLSFFPAMVHVASLCPMRSERYELFNMAVQOVHANINENSMEEVWR 356
 DB 275 -----ALVSTAILLVCLALLLPANSE-----IHGVLSTIF-----WG 309
 QY 357 IEMVTSFGMSGLSL-----LAVTSIPSVN-----ALMNRRESFIQST 397
 DB 310 IAMMIIQLGQVKKVLAAPATDVAMALFSGIENFIGNIGAGALVGNVSLHMS-----MSM 364
 QY 398 LGVYALISTFHVLIYG-----WKRAFEER 422
 DB 365 IGYVG-TVPAPALIMSIILFRMPVLEED 394

RESULT 11
 SOTB_ECOLI STANDARD; PRT; 396 AA.
 AC P31122; P77353; P76883;

DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Sugar efflux transporter.
 CN SOTB OR B1528.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_Taxid=562;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatter F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.
 RA "The complete genome sequence of Escherichia coli K-12.";
 RA Science 277:1453-1474(1997).
 RL (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97251357; PubMed=9097039;
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Itoh T.,
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Maki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
 RA Sankel G., Seki Y., Sivasubramanian S., Tagami H., Takeda J.,
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.,
 RA "A 570-kb DNA sequence of the Escherichia coli K-12 genome
 RA corresponding to the 28.0-40.1 min region on the linkage map.";
 RA DNA Res. 3:363-377(1996).
 RL (3)
 RP SEQUENCE OF 217-396 FROM N.A.
 RX MEDLINE=93186717; PubMed=8383113;
 RA Cohen S.P., Haechler H., Levy S.B.,
 RA "Genetic and functional analysis of the multiple antibiotic
 RA resistance (mar) locus in Escherichia coli.";
 RA J. Bacteriol. 175:1484-1492(1993).
 RL (4)
 RP CHARACTERIZATION.
 RC STRAIN=SR0;
 RX MEDLINE=99194728; PubMed=10094697;
 RA Bost S., Silva F., Belin D.;
 RA "Transcriptional activation of ydeA, which encodes a member of the
 RA major facilitator superfamily, interferes with arabinose accumulation
 RA and induction of the Escherichia coli arabinose PBAD promoter.";
 RA J. Bacteriol. 181:2185-2191(1999).
 RL (5)
 RP CHARACTERIZATION.
 RC STRAIN=JS219;
 RX MEDLINE=99369894; PubMed=10438792;
 RA Carole S., Pichoff S., Bouché J.-P.;
 RA "Escherichia coli gene ydeA encodes a major facilitator pump which
 RA exports L-arabinose and isopropyl-beta-D-thiogalactopyranoside.";
 RA J. Bacteriol. 181:5123-5125(1999).
 RL -1- FUNCTION: Involved in the efflux of sugars. The physiological role
 CC may be the reduction of the intracellular concentration of toxic
 CC sugars or sugar metabolites. Transports L-arabinose and to a
 CC lesser extent IPTG. Seems to contribute to the control of the
 CC arabinose regulation.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (Probable).
 CC -1- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN AS
 CC THE DRUG RESISTANCE TRANSLOCASE FAMILY). SOTB (TC 2.A.1.2)
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@sib-slb.ch.

CC EMBL; AE000250; AAC74601.1; -
 DR EMBL; D90795; BAA15210.1; -
 DR EMBL; D90796; BAA15218.1; -
 DR EMBL; D90797; BAA15230.1; -
 DR EMBL; M96235; -; NOT_ANNOTATED_CDS.
 DR Ecogen; EG11636; scb.
 DR InterPro; IPR003662; sub.transporter.
 DR Pfam; PF00083; sugar.tri.1.
 DR Transport; Sugar transporter; Transmembrane; Inner membrane;
 KM Complete proteome.
 FT TRANSMEM 15 35 POTENTIAL.
 FT TRANSMEM 50 70 POTENTIAL.
 FT TRANSMEM 81 101 POTENTIAL.
 FT TRANSMEM 103 123 POTENTIAL.
 FT TRANSMEM 136 156 POTENTIAL.
 FT TRANSMEM 170 190 POTENTIAL.
 FT TRANSMEM 209 229 POTENTIAL.
 FT TRANSMEM 246 266 POTENTIAL.
 FT TRANSMEM 275 295 POTENTIAL.
 FT TRANSMEM 299 319 POTENTIAL.
 FT TRANSMEM 333 353 POTENTIAL.
 FT TRANSMEM 364 384 POTENTIAL.
 SQ SEQUENCE 396 AA; 42538 MW; CB6A34CA4EE6D4F0 CRC64;

Query Match 4.1%; Score 104; DB 1; Length 396;
 Best local Similarity 18.8%; Pred. No. 2.7;
 Matches 85; Conservative 72; Mismatches 148; Indels 146; Gaps 20;

QY 31 VVVGVGSGDGFASKLIRICGHHVIGSRNPKFSEFPFHVVDVTHHEDATKTNITF 90
 DB 31 VVVGGLL-SDIASFHMQAQVCMITITAVVVALMSLPFMTLSOVERKLLICLFVVF 88
 QY 91 VAHREHYTSLMDRLHLLVGLKILIDVSNMRLINQPESSNAEYLASEPDSLIVKGFNVVS 150
 DB 89 IASHVLSFLS-WSFVTLVLSRI-----GVAFNAHIF-----WSTIA 123
 QY 151 ANALOGLP--KQASROYVICSNNIOAROVIELAR-----QLNFIPIIDGLS----- 195
 DB 124 SLAIRMAPAGKRAQALSLATGATMALVGLPGRIVGOYFGWRMFFAIGALITLLC 183
 QY 196 -----SAREIENPLRLFTLMRGPVVAISLAT-----PFELYSFQVDIHPY 229
 DB 184 LKILPLLSEHSGLSKSLPL--LFRPALMSITLLVYVVTATYATYST-----IEEP 235
 QY 240 ARN---QOSDFYKPIPIEIVNKTLPVATVLLSLVYLAGLLAAVYOLYGTXYRRFPWLE 296
 DB 236 VQNIAGSFNF-----ATALLLLGGAGIIGVIFGKLGQVNS----- 274
 QY 297 TWLOCRKQGLISFFPAMVAVASLCLPMRBSRYFLNMAVQOVHANIENSNNEEVR 356
 DB 275 -----ALVSTALALLVCLALLPANSE-----IHLGVLSIF-----WG 309
 QY 357 IEMVIFSGIMSLGLSL-----LAVTSIPSVSN-----ALNWRFSFIQST 397
 DB 310 IAMMIIIGLQKQVKTALAPDADVANAALFSGIFNIGIGALVGNQVSLHWS-----MSM 364
 QY 398 LGVALLISTFHVLYG-----WKRPFEE 422
 DB 365 IGVVG-AYPAFAALMSIIFRFRVPTLEEQ 394

RESULT 12
 Y049_AQUAE
 ID Y049_AQUAE STANDARD; PRT; 268 AA.
 AC 067364;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein AQ_1349.
 GN AQ_1349.

OS Aquifex aeolicus.
 OC Bacteria; Aquificales; Aquificaceae; Aquifex.
 NCBI_TaxID=63363;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VF5;
 RX MEDLINE=98196666; PubMed=9537320;
 RA Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 aeolicus";
 RL Nature 392:353-358(1998).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-slb.ch).

CC EMBL; AE000736; AAC07335.1; -
 DR Hypothetical protein; Transmembrane; Complete proteome.
 KM Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 5 24 POTENTIAL.
 FT TRANSMEM 37 59 POTENTIAL.
 FT TRANSMEM 69 91 POTENTIAL.
 FT TRANSMEM 104 126 POTENTIAL.
 FT TRANSMEM 130 152 POTENTIAL.
 FT TRANSMEM 173 195 POTENTIAL.
 FT TRANSMEM 210 232 POTENTIAL.
 FT TRANSMEM 245 267 POTENTIAL.
 SQ SEQUENCE 268 AA; 30387 MW; B3EA501514FPA3 CRC64;

Query Match 4.1%; Score 103.5; DB 1; Length 268;
 Best local Similarity 25.1%; Pred. No. 1.9;
 Matches 61; Conservative 41; Mismatches 70; Indels 71; Gaps 13;

QY 222 LATFFFLYS-FVNDVHPARNQOSDFY---KPIEIV-NKTLPIVATVLLSVYLAGLL 276
 DB 6 IALFWVIFSXYVDVL-----FFREAKIEREFCKNKL--ALSIVAGYFGLA 52
 QY 277 AAYQOLY-YCTKRRRPPWLETWLOCRKQGLISFFPAMVAVASLCLPMRBSRYFLN 335
 DB 53 FSFYSVFYFESLFE-----VYLIFVSFTLLGV-----YIEDL 87
 QY 336 MAYQOVHANIENSNNEEVRIMYISFGIMSLSLAVTSIPSVSNALNREFSFIQ 395
 DB 88 IFLRKIDL-----KEILR-----GNAGAGITOGIYFLSLGLISLSPFRKESFIL 133
 QY 396 S-----TLGVALLISTFHVLYIGMKRAPEEYRYFTPPNFVALVPSIVILGKI 447
 DB 134 SVIYSLIYISLGMVLFISTL-LMSRLKLNFEEYK-----ENFSASIVIGSTILGVSV 188
 QY 448 ILF 450
 DB 189 VLY 191

RESULT 13
 NP1_MOUSE
 ID NP1_MOUSE STANDARD; PRT; 465 AA.
 AC 061983;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Renal sodium-dependent phosphate transporter protein 1 (Sodium/phosphate
 DE cotransporter 1) (Na(+)/P1 cotransporter 1) (Renal sodium phosphate
 DE cotransporter 1) (Renal Na(+)-dependent phosphate cotransporter
 DE 1).
 GN SLC17A1 OR NP1.


```

QY 303 KOLGLISFFPAMVHVAISICLPMRSERYL-----ELNNAYQOVHANIENSANE--- 351
DB 143 -----FREDDITINLCSTNLTMYTILSLACCALDPLMMY-----TWNDRKK 184
QY 352 -----EEVWRIEMTISFIMISGLISLLAVISIPVS-----NALMW 388
DB 185 INVYDINSOLYKVEQHSHSLMAVSLNSLLQLEFVSIVAFISFVLAMNSMTNPELGNANMLRW 244
QY 389 -----REFSFIOST--LGYVALLISFPHVLIYGMKRAFEERYR---FTTPMFVL- 434
DB 245 FFAFKVLYRVSFHHSTHOLLFMSTIYADFHAIISGELOAI--SLRVSVPYDDPVMINE 302
QY 435 -----ALVPSIY-----ILKILFLPCISRLKRIKK-GWEKSOP- 470
DB 303 TGECAITWYGAKKMLMSFIQTALTIRIIRIIPVY-----KIKPKRQGLFNARA 358
QY 471 LEEGIGGT 479
DB 359 LTAGISTTI 367

RESULT 15
FSHR_PIG
AC P49059; 077514; STANDARD; PRT; 695 AA.
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Follicle stimulating hormone receptor precursor (FSH-R) (Follictrophin
DE receptor).
GN FSHR.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Ovary;
RX MEDLINE=96011644; PubMed=7590277;
RA Remy J.J., Labhib-Mansais Y., Yerie M., Bozon V., Couture L.,
RA Pajot E., Grebert D., Salese R.;
RT "The porcine follictrophin receptor: cDNA cloning, functional
RT expression and chromosomal localization of the gene.";
RL Gene 163:257-261(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-Ovary;
RA Wang Y.F., Meyer K.B., Schmidt K., Wan S.J., Degen S.J.F.,
RA la Barbera A.R.;
RT "Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.
RT - FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY
RT OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
RT ADENYLATE CYCLASE.
RT - SUBCELLULAR LOCATION: Integral membrane protein.
RT - SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
RT FSH/LSH/TSH SUBFAMILY
CC -1- SIMILARITY: CONTAINS 6 LEUCINE-RICH REPEATS (LRR).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L31966; AAA86933.1; -.
CC EMBL: AF025377; AAC24981.1; -.
CC HSSP: P23945; 1XUN.
CC GCRDB: GCR_1561; -.
CC InterPro: IPR000276; GPCR_Rhodpsn.
CC InterPro: IPR001611; LRR.

```

```

DR InterPro: IPR000372; LRR_Nterm.
DR Pfam: PF00001; 7tm_1; 1.
DR Pfam: PF00560; LRR_4; 4.
DR Pfam: PF01462; LRRNT; 1.
DR PRINTS: PRO0373; GLYCHROMER.
DR PRINTS: PRO1143; FSHRECEPTOR.
DR SMART: SM00013; LRRNT; 1.
DR PROSITE: PS00237; G_PROTEIN_RECP_F1_1; 1.
DR PROSITE: PS50262; G_PROTEIN_RECP_F1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW phosphorylation; Repeat; Leucine-rich repeat.
FT SIGNAL 1 17
FT CHAIN 18 695
FT DOMAIN 18 366
FT TRANSMEM 367 387
FT DOMAIN 388 398
FT TRANSMEM 399 421
FT DOMAIN 422 443
FT TRANSMEM 444 465
FT DOMAIN 466 485
FT TRANSMEM 486 508
FT DOMAIN 509 528
FT TRANSMEM 529 550
FT DOMAIN 551 573
FT TRANSMEM 574 597
FT DOMAIN 598 608
FT TRANSMEM 609 630
FT DOMAIN 631 695
FT REPEAT 631 68
FT REPEAT 69 93
FT REPEAT 119 143
FT REPEAT 170 192
FT REPEAT 193 216
FT REPEAT 218 240
FT DISULFID 442 517
FT CARBOHYD 191 191
FT CARBOHYD 199 199
FT CARBOHYD 293 293
FT CONFLICT 2 2
FT CONFLICT 13 13
FT CONFLICT 60 60
FT CONFLICT 166 166
FT CONFLICT 215 215
FT CONFLICT 247 247
FT CONFLICT 257 257
FT CONFLICT 334 334
FT CONFLICT 349 349
FT CONFLICT 352 352
FT CONFLICT 383 383
FT CONFLICT 407 407
FT CONFLICT 421 421
FT CONFLICT 427 427
FT CONFLICT 435 435
FT CONFLICT 483 483
FT CONFLICT 550 550
FT CONFLICT 586 586
FT CONFLICT 607 607
FT CONFLICT 691 691
SQ SEQUENCE 695 AA; 78172 MW; E9EBDB29C79C450 CRC64;

```

```

Query Match 4.1%; Score 103; DB 1; Length 695;
Best Local Similarity 16.0%; Pred. No. 6.1; Indels 212; Gaps 22;
Matches 91; Conservative 95; Mismatches 169;
18 LPNG-INGIKDARKVTVGVIGSGDFAKSLTIRLCRGYHVYIGSRNPKRASEFFPHVVDV 76
DB 61 IRKGAVSFGDLEKI-----EISQNDVLEVTIAN--VFSNPKL----- 97
QY 77 THHEDALTKNTIIFVAIHEHYTSLMDLRHLVG-----KLLDIVSN 119
DB 98 --HEIRIERANNML-LYIDPDACONLPLRYLLISNTGVKHLPAVHKIQSLQKVLADIODN 154

```

Oy	120	MRINQYPSNMEYLASEPDSLIIYKGFNVSAAMLOCPDASQOYICSNNTIQAQOYI	179
Db	155	INIH-----IYERNSEFVGLSEESMTLW-----SKNGIREIHNCA-----	189
Oy	180	ELAROLNPIPIDGLSSAREIENLPLRLFTIMRGPVVAISLATEFFFLYFVADYIHYP	239
Db	190	-----FNCTOLDLNLSDNDELLEELPNDVDFOGAGSPVILISRTIRLSLPSYGLENIKTL	244
Oy	240	ARNQOSDYKTRP-----IEIVNKTU-----PHVATLTL-----	267
Db	245	RAKSTYINLAKTRPLEKEFYTLMEASLYTPSHCCAFANMRROIIDLHPKCNKSLTROEYDVA	304
Oy	268	-----SLVYLGLLAAYQOLYGTGYRRF-----PRLWETWLOCRKOLG----	306
Db	305	TQARGQVSLADEDESSLAKFEPTMYSEFDYLCNEVYDICSPEPTFPCEDINGHDI	364
Oy	307	-----LLSFFPAMHVA-----YSLCLPMBRSEBTLFLNMAIQQVHANIENTSUNE	351
Db	365	LRLVLTWTSILATIGNIIVLILTSQYKTLVP-----RELMCNLAFAD-----	408
Oy	352	EEWVRIEWYISFGINSJGLSILSLAVTSIPSYSNALNRE-----FSFIOTSLG----	399
Db	409	-----LCIGIYLLLLIASVDLHTKQYHNHYAIDMOGACGDAGFFVFPASELSVTI	459
Oy	400	YVALLISTFH-----VLTYGKRPAEEERYRYETTPNEV-LALVLP	439
Db	460	LTATILEMHTITHAMOLCKVOLRHAASIMLVGMIFALVFALPIRGISSYKVASICLP	519
Oy	440	-----SLVILGKIILFLPC	453
Db	520	MDIDSPQLSYVSVLSLVNLNLAFAVYC	546

Search completed: May 9, 2002, 03:23:06
Job time: 4216 sec

Db 64 TAGLPSLAQVTFQBEAVSPEVIFVAVFRHHYSLCSLADQAGKILVDVSNPTKEERL 123
 QY 124 QYPSNAEYLAFLPPDLIVKGFVNVSAMALQLPKDSAROVYICSNINQAROOVIELAR 183
 Db 124 QHRQSNMEYLAFLPACVIVKAFNVISAMALQACRQDNROVLLCGDLEAKHYSSEAR 183
 QY 184 QLANFIPDLGSLSSAREIENLPLRLFTLMGRPVVAISLATFFLYSFVRDVIHPYARNO 243
 Db 184 AMGTFPLDMGSLASAREVEAIPRLPLPSMKVPTLLALGSLQSYAAYNIRDVLPYIRKD 243
 QY 244 OSDPYKPIELIVNKTPIVATITLSLVYAGLAAAOVLYGTYKRRPPELMTWLOCRK 303
 Db 244 ENKRYKAPLISVNTTLPVAVVLLSLVYLPGLVLAALQLRGRTKYQRPDMLDHMLDRK 303
 QY 304 QLGILSFEFAMVHVAYSICLPMRSEERYLFNMAVQOVHANIENSNNEEVWRIEMYSF 363
 Db 304 QIGLISFEFAMVHVAYSICLPMRSHRDYLVNLAQVLAANKSRMLWEEVWRIEYLSL 363
 QY 364 GIMSLGILSLAVSIPSVSNAALMWRSEFIOSTGLVALLISPHVLYGKRAFEERY 423
 Db 364 GVLAMGLSLAVSIPSISANSLMWKSEFVOSTLGEVALLSTMHTLTLYGWTAFEBNH 423
 QY 424 YRFYTPPFEVLAFLVPSVILGKIILFLPCISRLKRIKKGMEK 467
 Db 424 YKFYLPPTFTLLPVCYIILAKGLFLPLCLSHRLTKIRGWER 467

RESULT 2
 ID Q92421 PRELIMINARY; PRT: 514 AA.
 AC Q92421:
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE DUDULIN 2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Serru V., Lambdin D., Lenoir C., Manivet P., Vaudourdolle M.,
 RA Kellermann O., Loric S.;
 RT "Molecular cloning and expression of mouse dudulin 2."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY029586; AAK50539.1;
 SQ SEQUENCE 514 AA; 57268 MW; 3398866C28BAEC02 CRC64;

Query Match 54.5%; Score 1382; DB 11; Length 514;
 Best Local Similarity 56.0%; Pred. No. 7.4e-101;
 Matches 260; Conservative 87; Mismatches 105; Indels 12; Gaps 4;

QY 6 MMGSPKSLSEETCLNGINGIKDARKVYGVYSGDFAKSLTRLIRCGYHNVIGSRNPF 65
 Db 14 LVDSGSLAE--VP-----EAPK--VGLISGDFARSLATRLVSGSFVYVSGSRNPKR 63
 QY 66 ASEFPFVVDVTHHEDALTNTNIIFVAIHREHYSLMDLRLLVGLKILIDVSNMNRIN-- 123
 Db 64 TAGLPSLAQVTFQBEAVSPEVIFVAVFRHHYSLCSLADQAGKILVDVSNPTKEERL 123
 QY 124 QYPSNAEYLAFLPPDLIVKGFVNVSAMALQLPKDSAROVYICSNINQAROOVIELAR 183
 Db 124 QHRQSNMEYLAFLPACVIVKAFNVISAMALQACRQDNROVLLCGDLEAKHYSSEAR 183
 QY 184 QLANFIPDLGSLSSAREIENLPLRLFTLMGRPVVAISLATFFLYSFVRDVIHPYARNO 243
 Db 184 AMGTFPLDMGSLASAREVEAIPRLPLPSMKVPTLLALGSLQSYAAYNIRDVLPYIRKD 243
 QY 244 OSDPYKPIELIVNKTPIVATITLSLVYAGLAAAOVLYGTYKRRPPELMTWLOCRK 303
 Db 244 ENKRYKAPLISVNTTLPVAVVLLSLVYLPGLVLAALQLRGRTKYQRPDMLDHMLDRK 303

QY 304 QLGILSFEFAMVHVAYSICLPMRSEERYLFNMAVQOVHANIENSNNEEVWRIEMYSF 363
 Db 304 QIGLISFEFAMVHVAYSICLPMRSHRDYLVNLAQVLAANKSRMLWEEVWRIEYLSL 363
 QY 364 GIMSLGILSLAVSIPSVSNAALMWRSEFIOSTGLVALLISPHVLYGKRAFEERY 423
 Db 364 GVLAMGLSLAVSIPSISANSLMWKSEFVOSTLGEVALLSTMHTLTLYGWTAFEBNH 423
 QY 424 YRFYTPPFEVLAFLVPSVILGKIILFLPCISRLKRIKKGMEK 467
 Db 424 YKFYLPPTFTLLPVCYIILAKGLFLPLCLSHRLTKIRGWER 467

RESULT 3
 ID Q9NVB5 PRELIMINARY; PRT: 488 AA.
 AC Q9NVB5:
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE CDNA FLJ10829 FIS, CLONE NT2RP4001138 (DUDULIN 2).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Aotaka S., Yoshikawa Y.,
 RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
 RA Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;
 RT "NDO human cDNA sequencing project."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Serru V., Manivet P., Lenoir C., Eschwege P., Lambdin D.,
 RA Vaudourdolle M., Kellermann O., Loric S.;
 RT "Dudulin 2, a new tumor antigen expressed in various human tumors."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK001691; BA091839.1;
 DR EMBL: AY029585; AAK50538.1;
 DR InterPro: IPR003006; IG_MHC.
 DR PROSITE: PS00290; IG_MHC; UNKNOWN_1.
 SQ SEQUENCE 488 AA; 54616 MW; BC0BCA48335AAD6 CRC64;

Query Match 53.5%; Score 1356; DB 4; Length 488;
 Best Local Similarity 53.1%; Pred. No. 7.8e-99;
 Matches 260; Conservative 91; Mismatches 121; Indels 18; Gaps 6;

QY 3 SISMMGSPKSLSEETCLNGINGIKDARKVYGVYSGDFAKSLTRLIRCGYHNVIGSRN 62
 Db 11 SLHLVDSGSLAE--VP-----EAPK--VGLISGDFARSLATRLVSGSFVYVSGSRN 60
 QY 63 PKFASEFPFVVDVTHHEDALTNTNIIFVAIHREHYSLMDLRLLVGLKILIDVSN--NM 120
 Db 61 PKRTARLYPSSAQAQVTFQBEAVSPEVIFVAVFRHHYSLCSLSDQAGKILVDVSNPTKE 120
 QY 121 RINQYPSNAEYLAFLPPDLIVKGFVNVSAMALQLPKDSAROVYICSNINQAROOVIE 180
 Db 121 EHLQHRQSNMEYLAFLPACVIVKAFNVISAMALQACRQDNROVLLCGDLEAKHYSSEAR 180
 QY 181 QLANFIPDLGSLSSAREIENLPLRLFTLMGRPVVAISLATFFLYSFVRDVIHPYARNO 240
 Db 181 MALAMGFMVDMGSLASAREVEAIPRLPLPSMKVPTLLALGSLQSYAAYNIRDVLPYIR 240
 QY 241 RNOOSDFYKPIELIVNKTPIVATITLSLVYAGLAAAOVLYGTYKRRPPELMTWLOCR 300
 Db 241 QESQNKKEFKPLISVNTTLPVAVVLLSLVYLPGLVLAALQLRGRTKYQRPDMLDHML 300
 QY 301 CRKQGLSFEFAMVHVAYSICLPMRSEERYLFNMAVQOVHANIENSNNEEVWRIEMYS 360
 Db 301 HRKQIGLISFEFAMVHVAYSICLPMRSHRDYLVNLAQVLAANKSRMLWEEVWRIEYLS 360

[illegible]

ID	Q91W31	PRELIMINARY;	PRT;	470 AA.
AC	Q91W31;			
DT	01-DEC-2001 (TREMBLrel, 19, Created)			
DT	01-DEC-2001 (TREMBLrel, 19, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel, 19, Last annotation update)			
DE	TRANSMEMBRANE PROTEIN TIARP.			
OS	TIARP.			
CS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=SWISS; TISSUE=ADIPOSE TISSUE;			
RX	MEDLINE=21424005; Pubmed=11443137;			
RA	Moldes M., Lasnier F., Gauthereau X., Klein C., Patraut J., Feve B.,			
RA	Chambaut-Guein A.M.;			
RT	"Tumor necrosis factor-alpha-induced Adipose-related protein (TIARP),			
RT	a cell-surface protein that is highly induced by tumor necrosis			
RT	factor-alpha and adipose conversion.";			
RL	J. Biol. Chem. 276:33938-33946(2001).			
DR	EMBL; AJ319746; CAC41351.1; --			
SO	SEQUENCE 470 AA; 52971 MW; F19239188E08663E CRC64;			

Query Match	44.8%;	Score 1134;	DB 11;	Length 470;
Best Local Similarity	47.5%;	Pred. No. 2.4e-81;		
Matches 212;	Conservative 90;	Mismatches 142;	Indels 2;	Gaps 2

QY 33 VGVIGSGDFAKSLTIRLCRGYHVVIGSRNPKFASEFFPHVDVYTHHEDALTKTNIFVA 92
| : ||| | : : : | : ||| : | : : : ||| : |
Db 22 VCIEGTGDFEKSGLGMLQCGSIVFGRNPQ-VSSLLPRGAETVSYSEAKSKSDIILA 80

Query Match	45.0%	Score 1141	DB 11	Length 470
Best Local Similarity	48.0%	Pred. No. 6.7e-82		
Matches 214	Conservative 89	Mismatches 141	Indels 2	Gaps 2
Qy	33	VGVISSGDPFASLTIRLLRCGHVAVVIGSNBPKFASFPFHVVVDYTHHEDALTKTNIIFVA	92	
Db	22	VCIFSTGPGFSGSLGKMLQCGSTIVFGSNRPQ-VESLLPRGAELVLSYSPAASKSDITIIIA	80	
Qy	93	IHRHRYTSLAMDRLHLVAKKILIDVSNMNMKINQYPSNMAEYASLAFPSDSLIVGKGVNVA	152	
Db	81	MHREHYDSELTETLYDLKKGVLVDVSNKKRKINQYPSNMAEYASLQLEBGAVVAVAFNTISW	140	
Qy	153	ALQGPDKASROYVICSNNIQARQOVIETLARQNFPIIDQSLSARELENTPLRLFTTM	212	
Db	141	ALQSTLIDASRQVFCGDNDSKAKQOVMQIDARLGLPIIDQSLMASELENTPLQLEPVM	200	
Qy	213	RGPVVAIVSLATFEFLYSEVRDVIHPYARNQOQDPYKIPDIETVKNKTLPIVAITLLSLVY	272	
Db	201	RPFPYLSVLCVCFEYFVYCAIREVIRPYVNGKDAYETRAISIPNRFPITALITLLAVYL	260	
Qy	273	AGLLAAAOVLYGCTYRRRPPWLEWLOCRKQGLISFFPAMVHVASICLPMRSERL	332	
Db	261	PGILLAIITLLIRGKYRRFPNWLDMMLCRKQGLVAGLGAFAFLVYTLVPIRYYVRR	320	
Qy	333	FLNMAVQOVHANIENSMNEEYWRLEMYISFCIMSGILSLTSLAVTSIPVSNAIIMRERS	392	
Db	321	LRNATITDALINKSDPFTTSTAMINDSTIALDGLIDFELLGLIGTISLPSVSNMWNREPR	380	
Qy	393	FIQSTLGYVALLISTFHVLIYGWKRAFEVEEYRYETPPNFVALVLPISVILGKITLLP	452	
Db	381	FVQSKGLGYTLVLTAFHTLVYGGKRFLLPSILRMSLPSAYILAVIPCAVLVLCILIMP	440	
Qy	453	CISRKLRATIKKGWKR-SQFLBEGITG	477	
Db	441	CIDKLTETRIQGWERNISKYTQSALNG	466	

Qy	153	ALOGPKASQOVYICGSNNIARQOVIEIARNOLEFIPIDIGSSISAREIENLPLRPTLM	212
Qy	141	ALOGSTIDASQOVYEVCGSDSKAKHRMVDIATLTLPLPDGOSLMAASEIENYPLQLEPMW	200
Qy	213	RGPVVAIATLAEFFELVSFVADVTHHPARNOQSDPFYKIPRIEIVNKTLPYATITLSTLYL	272
Dd	201	RFPFLTSVLCIIEFFVYCAIEVYIYYPVNGKTDITATYLAISIPRRVPIITALLILALVYL	260
Qy	273	AGLLAAAYQLYYGTKRFRFPWMLTLOCRKQGLLSLFFPAMVHAYASLCLPMKRSERYL	332
Dd	261	PGLIAAILQLYRGTKRYRRFPWMLDHMMLCRKQGLGVALGAFELHVIATVITPIRYVYWR	320
Qy	333	FLNMAAYQOVVHANIENSNNEEVWPIEMTISGIMSGLSILSLAVTSTPVSNALNMRFS	352
Dd	321	LRNATITQALNNKSPFITTSYAMINDSYLAGILGIFELFLTGITTSIPVSNMWNREFR	380
Qy	393	FIOSTLGVALLISPHLYIYGMKRAFEDEEYRFPYPPNVIALVPSYILKILFLP	452
Dd	361	FVQSLGTLVTLCTAHTLTVGGKRELPSPSTLRWSDPSATYIALITPCAVLVLCJILMP	440
Qy	453	CISRKLRKIKKGMKE-SQLEEGIG 477	
Dd	441	CIDKTLTRIRQGMERNKRYTQALING 466	
RESULT	6		
Q91ZB8			
ID	Q91ZB8	PRELIMINARY;	PRT; 474 AA.
AC	Q91ZB8		
DT	01-DEC-2001	(TIREMBLrel, 19, Created)	
DT	01-DEC-2001	(TIREMBLrel, 19, last sequence update)	
DT	01-DEC-2001	(TIREMBLrel, 19, last annotation update)	

RESULT
Q91W31

OC Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus

OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Serru V., Lambdin D., Manivet P., pernet P., Vaubourdolle M.,
 RT Kellermann O., Loric S.,
 RT "Molecular cloning and expression of two new members of the dudulin
 family".
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY029778; AAK40270.1;
 SO SEQUENCE 474 AA; 53543 MW; 937EBA3393D13956 CRC64;

Query Match 43.2%; Score 1094.5; DB 11; Length 474;
 Best Local Similarity 47.8%; Pred. No. 3.2e-78;
 Matches 208; Conservative 83; Mismatches 143; Indels 1; Gaps 1;

QY 33 VGVIGSDGDFAKSLTIRLCRGYHYVIGSRNPKFSEFPFHVVDYTHEDALTKNTIFVA 92
 DB 22 VCFGTGDFGSKGMKMLQCGYSIVFGSRNPQ-VSSLLPRCAEVLYSSEAKSKSDIILA 80
 QY 93 IHRHHTSLMDLRHLVGLKILIDVSNMNRINOPESSNAEYLALEPDSLIYKGFNVVSAM 152
 DB 81 MHRHHTSLMDLRHLVGLKILIDVSNMNRINOPESSNAEYLALEPDSLIYKGFNVVSAM 140
 QY 153 ALQGPDAARQYVIGSNNTQARQVIELARQNLFPIDIGLSSAREIENPLRFTLW 212
 DB 141 ALQGPDAARQYVIGSNNTQARQVIELARQNLFPIDIGLSSAREIENPLRFTLW 200
 QY 213 RGPVVAISLATEFFELSFVADVHPARNOQSDFKIPRIEIVNKTPIYATILSLVYL 272
 DB 201 RFPYLSLVLCIFELFCALREVYIPYMGNTDVTYRLAISTPKRVPITALLLAVYL 260
 QY 273 AGLLAAVQOLYGTGYRRFPFMEYTLQCRKQGLSLFFPAMVVAVSLCPMRSEFYS 332
 DB 261 PGILAILQYRGYKRYRPFMDHMLCKRQGLSLGAFALHAIYTVIPYVYRMR 320
 QY 333 FLNAAVQVHANINENSNEEVRRIEMTISFGIMSLGLSLAVTSIPSVSNALNREFS 392
 DB 321 LRKATITQALTSKROPITSTYAMINDSYLALGIFFLFLVIGTITSPSVSNMNMREFR 380
 QY 393 FIOSTIGVALLSTFPHVLYGKRAFEERYRFPNPLVLAIVLSIYLGIIIFLP 452
 DB 381 FVOSKLGIVLVCTANTLVYGGKRLSPSLKSLPSAVIALVYPCAVLVLCILMP 440
 QY 453 CISRKLRIKKGWEK 467
 DB 441 CIDKTLIRIRGGLK 455

RESULT 7
 Q9GL50 PRELIMINARY; PRT; 338 AA.
 AC Q9GL50;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DE SIX TRANSMEMBRANE ENDOTHELIAL ANTIGEN OF PAEC.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Nagasaki T., Boulay G., Couplé S., Coulon F., Tesson L.,
 RA Heslan J.-M., Soullion J.-P., Charreau B.;
 RT "Differential gene expression in endothelial cells during TNF- α -
 RT and LPS-mediated activation".
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF319659; AAG33868.1;
 KW Transmembrane.
 SO SEQUENCE 338 AA; 39918 MW; ED490E86E067A32B CRC64;

Query Match 31.6%; Score 801; DB 6; Length 338;
 Best Local Similarity 49.7%; Pred. No. 2.9e-55;
 Matches 148; Conservative 64; Mismatches 78; Indels 8; Gaps 2;

QY 170 NNTARQVIELARQNLFPIDL-GISLSAREIENLRLFTLMRGVVAISLATEFFL 228
 DB 34 NSMRRPMLVHLQOTAFHDEFDCPELQHOE-----LEPKWHLPIKIALVSSLTFL 86
 QY 229 YSFVADVHPARNOQSDFKIPRIEIVNKTPIYATILSLVLAAGLLAAVQOLYGTGY 288
 DB 87 YTLREVIHPFVTSHQOYFKIPILVINKVLPKMSITLLALVLPYVIAIVQAHNTKTX 146
 QY 289 RRPFWLETLQCRKQGLSLFFPAMVVAVSLCPMRSEERYELNMAVQVHANINENS 348
 DB 147 KRPFWLMDRMVTRKQFGLSLFFPAMVVAVSLCPMRSEERYELNMAVQVQOONKEDA 206
 QY 349 WNEEVRRIEMTISFGIMSLGLSLAVTSIPSVSNALNREFSIQSTIGYVALLSTF 408
 DB 207 WIEDVVRMEIYVSLGIVTALIALAVTSIPSVSDSLWREFFHYIOSTIGIYSLGTTI 266
 QY 409 HVLVIGKRAFEERYRFPNPLVLAIVLSIYLGIIIFLPCISRLKRIKKGWE 466
 DB 267 HALIFAMNKVNDIKQFTWTPPTFMIAVFLPYVLYICKVILLPLCLRKILKIRHGE 324

RESULT 8
 Q92422 PRELIMINARY; PRT; 339 AA.
 AC Q92422;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DE DDDULIN.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Serru V., Manivet P., Lambdin D., Vaubourdolle M., Kellermann O.,
 RA Loric S.;
 RT "Prostate and non-prostate expression of dudulin, the mouse ortholog
 RT of human STRAP".
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY029584; AAK50537.1;
 SO SEQUENCE 339 AA; 39109 MW; 32A2C29FE2E33BD0 CRC64;

Query Match 30.3%; Score 768; DB 11; Length 339;
 Best Local Similarity 54.7%; Pred. No. 1.2e-52;
 Matches 141; Conservative 53; Mismatches 64; Indels 0; Gaps 0;

QY 209 FTLMRGVVAISLATEFFLVSFVADVHPARNOQSDFKIPRIEIVNKTPIYATILSL 268
 DB 68 FPMNRPLPVKVAATISSTLYTLREITLPLVTSKREYFKIPILVNTTLPVAVYLLS 127
 QY 269 LVYLAGLLAAVQOLYGTGYRRFPFMEYTLQCRKQGLSLFFPAMVVAVSLCPMRRS 328
 DB 128 LVYLAGLLAAVQOLYGTGYRRFPFMEYTLQCRKQGLSLFFPAMVVAVSLCPMRRS 187
 QY 329 ERYLFLMAVQVHANINENSNEEVRRIEMTISFGIMSLGLSLAVTSIPSVSNALN 388
 DB 188 HRVDLVMLAVKQVLANKEADVDEHDVVRMEIYVSLGIVTALIALAVTSIPSVSDSLTW 247
 QY 389 RERSFIQSTIGYVALLSTFPHVLYGKRAFEERYRFPNPLVLAIVLSIYLGIIIFLP 448
 DB 248 RERSFIQSTIGYVALLSTFPHVLYGKRAFEERYRFPNPLVLAIVLSIYLGIIIFLP 307
 QY 449 LFLPCISRLKRIKKGWE 466
 DB 308 LCLPCLRKILKIRGWE 325

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki K., Sato K., Schoenbach C., Segal T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wushaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL: AK015015; BAB29672.1; -
 DR MGD: MGI:1921301; 4921538B1.7Rik.
 SQ SEQUENCE 132 AA; 14931 MW; B56F8732AEBEAEF3 CRC64;

Query Match 25.3%; Score 642; DB 11; Length 132;
 Best Local Similarity 93.2%; Pred. No. 3e-43; Mismatches 0; Gaps 0;
 Matches 123; Conservative 6; Indels 0;

QY 359 MYISFGIMSLGLSLAVTSIPSVSNALNMRFSFIQSTIGYVALLSTFHVLYGKRA 418
 DB 1 MYISFGIMSLGLSLAVTSIPSVSNALNMRFSFIQSTIGYVALLSTFHVLYGKRA 60
 QY 419 FEERYFYPPNPNVLAIVPSIYILKILFLPCISRKLRKKGWKSQFLEEGIGT 478
 DB 61 FAERYFYPPNPNVLAIVPSIYILKILFLPCISRKLRKKGWKSQFLEEGIGT 120
 QY 479 IPHVSPEPVYV 490
 DB 121 VPHISPEPVYV 132

RESULT 12
 Q9H5R1 PRELIMINARY; PRT; 283 AA.
 AC Q9H5R1:
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE CDNA: FLJ23153 FIS, CLONE LNC09441.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LUNG;
 RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
 RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,
 RA Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Ota T., Suzuki Y.,
 RA Isogai T., Sugano S.;
 RT "MEDO human cDNA sequencing project.";
 RT Submitted (FUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK026806; BAB15559.1;
 DR InterPro: IPR000585; Hemopexin.
 DR PROSITE: PS00024; HEMOPEXIN; UNKNOWN_1.
 SQ SEQUENCE 283 AA; 31814 MW; DDD330DA7C607D06 CRC64;

Query Match 24.8%; Score 628; DB 4; Length 283;
 Best Local Similarity 43.2%; Pred. No. 1e-41;
 Matches 123; Conservative 56; Mismatches 80; Indels 26; Gaps 5;

QY 13 LSENC---LNGINGIDARKRVYGVSGDFAKSLIRLRRCGYHYVTSRNPKESEF 69
 DB 1 MEKTCIDALPLVMN--SSEKQETVCITGTDGFGSLGKMLQCGSYVFGSRNPQ-KTTL 57
 QY 70 FPHVVDTHEDALTKNIIFVAIHREHYTSLMDLRHLVGKILIDVSNMNRINOYPSN 129
 DB 58 LPSGAELVLSSEAKKDDIIIIAIHREHYDFELTELVLNGKILVDSNNLKNIOYPSN 117
 QY 130 AEYLASLPDLSLVKGFNVSAVALDGPXDASROYICSNNOARQOVIELARQLNFI 189

||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 DB 118 AEYLAVHVPGAHVVKANNTISAMALOGADASHQVFCVGNDSKAKORVMDIVNGLTTP 177
 QY 190 IDGSLSSARETELPRLETLIMRGPPVVAISLATFPELTSFVDRVHPYARNQDSFEK 249
 DB 178 MDQSLMAAKEIEKYPQLDLPMMRFPYLSVLCVLEFFCVLRDVIYPPYVEKKDNFR 237
 QY 250 IPRIEIVKTLPIVAITLLSLVYAGLAAAYQLYGTGRYRFPW 294
 DB 238 MAISIPNRIPI-----TAVYTACFG-----LPPW 262

RESULT 13
 Q9Y6U5 PRELIMINARY; PRT; 264 AA.
 ID Q9Y6U5:
 AC Q9Y6U5:
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE WGSF:H_RG087E15.1 PROTEIN (FRAGMENT).
 GN WGSF:H_RG087E15.1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE:99063792; PubMed=9847074;
 RA Sulston J.E., Waterston R.;
 RT "Toward a complete human genome sequence.";
 RL Genome Res. 8:1097-1108(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Strong C., Layman D., Graves T., Stromatt C.;
 RT "The sequence of Homo sapiens BAC clone CH8-87E15.";
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Waterston R.;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC005061; AAD43182.1; -
 FT NON_TER 1
 FT NON_TER 1
 SQ SEQUENCE 264 AA; 31203 MW; 35C948300357E72 CRC64;

Query Match 22.8%; Score 577; DB 4; Length 264;
 Best Local Similarity 56.4%; Pred. No. 9.8e-38;
 Matches 106; Conservative 36; Mismatches 46; Indels 0; Gaps 0;

QY 208 LFTLMRGPVVAISLATFPELTSFVDRVHPYARNQDSFEKIPRIEIVKTLPIVAITLL 267
 DB 77 LFPQMHPIKIAVMSLFLYTLRLREVHPPLATSHQOYFKIPILVINKVLPWVSITLL 136
 QY 268 SLVYLGLAAAYQLYGTGRYRFPWLETLWLCROKGLSFPFAMVHAYSLCPMR 327
 DB 137 ALVYLPVIAAYOVHNGTGYKKFPHWLDWMLTRKQFGLSFFAVLHAYTLSTAMRR 196
 QY 328 SERLPLFNMAVYOVHANIENSWNEEWEVRIEMTISFGIMSLGLSLAVTSIPSVSNALN 387
 DB 197 SYRKLLNMAVYOOVKEDAMIEHDVWRMEIYLSGLVLAITLAVTSIPSVSDSLT 256
 QY 388 WREFSFIO 395
 DB 257 WREFHYIO 264

RESULT 14
 Q9H7Y1 PRELIMINARY; PRT; 143 AA.
 ID Q9H7Y1:
 AC Q9H7Y1:
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

Query Match	6.3%;	Score 160;	DB 16;	Length 198;
Best Local Similarity	26.7%;	Pred. NO. 5.3e-05;		
Matches 55; Conservative	40;	Mismatches 83;	Indels 28;	Gaps 8

```
Search completed: May 9, 2002, 03:22:15
Job time: 4212 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2002, 22:28:40 ; Search time 308.42 Seconds
(without alignments)
10526.831 Million cell updates/sec

Title: US-09-802-520-2

Perfect score: 1891

Sequence: 1 ggggaagcagctgagtgctgcg.....gtcaattatcgtggttga 1891

Scoring table:

IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_032802:*

1:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24:	/SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1608.8	85.1	4329	22	AAS15801 Human ORF of Six-T
2	1534.8	81.2	1680	22	AAS15802 Human cDNA encodin
3	1466.2	77.5	1725	22	AAS15793 Human DNA for Six-
4	1409.8	74.6	2453	22	AAD07072 Human six transmem
5	1369.2	72.4	1561	22	AAS15811 Human cDNA encodin
6	1367.6	72.3	2238	22	AAS15810 Human ORF2 of Six-
7	1216.8	64.3	2102	22	AAS15813 Human cDNA encodin
8	1093.2	57.8	2192	23	AAS76493 DNA encoding novel
9	1093.2	57.8	3900	23	AAS64300 DNA encoding novel

10	530.6	28.1	1953	22	ABA45820 Human breast cell
11	530.6	28.1	1953	22	ABA56335 Human foetal liver
12	530.6	28.1	1953	22	ABA25976 Probe #4442 for ge
13	530.6	28.1	1953	22	AAK04514 Human brain expres
14	530.6	28.1	1953	22	AAK30016 Human bone marrow
15	530.6	28.1	1953	22	AAK14606 Probe #4539 for ge
16	530.6	28.1	1953	22	AAI35978 Probe #4664 used t
17	530.6	28.1	1953	22	AAI104420 Probe #4411 used t
18	528.6	28.0	539	22	ABA50938 Human breast cell
19	528.6	28.0	539	22	ABA68917 Human foetal liver
20	528.6	28.0	539	22	ABA35870 Probe #1436 for g
21	528.6	28.0	539	22	AAK17249 Human brain expres
22	528.6	28.0	539	22	AAK43038 Human bone marrow
23	528.6	28.0	539	22	AAI23801 Probe #13739 used
24	528.6	28.0	539	22	AAI14913 Probe #17799 used
25	528.6	28.0	539	22	AAI09410 Probe #9401 used t
26	528	27.9	528	22	AAI15797 Human Six-Transmem
27	528	27.9	528	22	AAI15806 Human Six-Transmem
28	523.4	27.7	525	22	AAI15796 Human Six-Transmem
29	523.4	27.7	525	22	AAI15805 Human Six-Transmem
30	519	27.4	519	21	AAZ49398 Human STRAP-2 part
31	476.4	25.2	1213	21	AAZ45296 Human prostate gro
32	425.8	22.5	444	19	AAV61348 Forward DNA sequen
33	425.8	22.5	444	19	AAV58684 Prostate tumour sp
34	425.8	22.5	444	21	AAA06447 Human immunogenic
35	425.8	22.5	444	22	AAI63655 Human prostate CDN
36	425.8	22.5	444	22	AAI10206 Human prostate tum
37	425.8	22.5	444	22	AAH93563 Human prostate-spe
38	425.8	22.5	444	22	AAH84877 Human prostate-tum
39	425.8	22.5	444	22	AAH02628 Rat p-HYDE coding
40	406.6	21.5	2714	22	AAH89167 AdSVphYHDE region
41	406.6	21.5	3885	22	AAH89169 Human foetal liver
42	395	20.9	395	22	ABA57377 Probe #5373 for ge
43	395	20.9	395	22	ABA26909 Human brain expres
44	395	20.9	395	22	AAK05412 Human bone marrow
45	395	20.9	395	22	AAK31011

ALIGNMENTS

RESULT	1
ID	AAS15801
ID	AAS15801 standard; cDNA; 4329 BP.
AC	AAS15801;
DT	16-JAN-2002 (first entry)
DE	Human ORF of Six-Transmembrane Protein of Prostate 1, STMP1.
XX	Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer
KW	benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW	cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW	leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW	liver cancer; lung cancer; cytostatic; ss.
XX	
OS	Homo sapiens.
XX	
Key	Location/Qualifiers
FT	178..1650
FT	/*tag= a
FT	/product= "STMP1"
XX	
PN	WO200172962-A2.
XX	
XX	04-OCT-2001.
PD	
XX	
PE	23-MAR-2001; 2001WO-US09410.
XX	
PR	24-MAR-2000; 2000US-191929P.
XX	
PA	(SAAT/) SAATCIOGLU F.

XX Saactloglu F:
PI WPI: 2001-662926/76.
XX P-PSDB: AAU10187.
DR
XX New polynucleotide for the diagnosis, prevention and treatment for
PT prostate and testis disorders, particularly prostate cancer, comprises
PT prostate-specific or testis-specific nucleic acids
XX
PS Claim 5: Fig 4D; 114pp: English.
XX
CC The invention relates to substantially pure prostate-specific or
CC testis-specific polypeptides and the nucleic acids encoding them.
CC Also included are vectors and host cells expressing the proteins, a
CC transgenic animal expressing the protein, antibodies against the
CC proteins, probes for detecting the nucleic acids, antisense molecules
CC for the nucleic acids and methods of isolating modulators of the
CC proteins. Compounds that modulate the prostate specific or testis
CC specific polypeptide are useful to diagnose, prevent or treat disorders
CC of the testis or prostate particularly prostate cancer, benign
CC prostatic hyperplasia, acute prostatitis, testicular cancer,
CC cryptorchidism, undescended, retractile, ascending or vanished
CC testis. Other proliferative disorders for which the modulators may be
CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC cancer, pancreatic cancer, liver cancer and lung cancer. The
CC present sequence represents the open reading frame of a prostate
CC specific protein, Six-Transmembrane Protein of Prostate 1, STMP1.
XX
SQ Sequence 4329 BP; 1315 A; 817 C; 790 G; 1407 T; 0 other:

Query Match 85.1%; Score 1608.8; DB 22: Length 4329;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1613; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 272 atattcaagaagatctctgtgatcttggaagtgccgtatcatggaataatctctat 331
DB 135 atttccagaagatctctgtgatcttggaagtgccgtatcatggaataatctctat 194
QY 332 gatgggaagccttaagagccttagtgaactgtttactctatgagcctaagtatcaa 391
DB 195 gatgggaagccttaagagccttagtgaactgtttactctatgagcctaagtatcaa 254
QY 392 agatgcagaagaagtcactgtaggtgtagtggaagtggaatttgcgaatccttgac 451
DB 255 agatgcagaagaagtcactgtaggtgtagtggaagtggaatttgcgaatccttgac 314
QY 452 catcgcactattagatgagcgtatcatgtagtgcataaggaagtagaattcctgaattgc 511
DB 315 catcgcactattagatgagcgtatcatgtagtgcataaggaagtagaattcctgaattgc 374
QY 512 ttctgaatttttctcattgtgtagatgtagtgcataaggaagtagaattcctcacaanaac 571
DB 375 ttctgaatttttctcattgtgtagatgtagtgcataaggaagtagaattcctcacaanaac 434
QY 572 aaaaataaatttctgctatatacagagaacattatacctccctgtggagcccgagaca 631
DB 435 aaaaataaatttctgctatatacagagaacattatacctccctgtggagcccgagaca 494
QY 632 tctctgtgtgggtaaaatcctgattgtagtgcataaacaatgagataaaccagatcac 691
DB 495 tctctgtgtgggtaaaatcctgattgtagtgcataaacaatgagataaaccagatcac 554
QY 692 agaatccaatgctgaataatttggctcattatccagatcttggatttgcagaagatt 751
DB 555 agaatccaatgctgaataatttggctcattatccagatcttggatttgcagaagatt 614
QY 752 taatgttctcagacttggcacttcagtttagagacctaagagatgcagccggcaggttta 811
DB 615 taatgttctcagacttggcacttcagtttagagacctaagagatgcagccggcaggttta 674
QY 812 tatatgcagcaacaataattcaagcgcaacaaggtattgaacttgcgcagattgaa 871

DB 675 tatatgcagcaacaataattcaagcgcaacaaggtattgaacttgcgcagattgaa 734
QY 872 ttcatcccatgacttgatgatactcctatcatcagccagaagaattgaaattaccct 931
DB 735 ttcatcccatgacttgatgatactcctatcatcagccagaagaattgaaattaccct 794
QY 932 aagactcttactctctgagagaagcgagtgtagtataaagcttgcgacatttt 991
DB 795 aagactcttactctctgagagaagcgagtgtagtataaagcttgcgacatttt 854
QY 992 ttcccttattcctttgtcagaagatgtagtattcatccatattgtcgaacccaagatga 1051
DB 855 ttcccttattcctttgtcagaagatgtagtattcatccatattgtcgaacccaagatga 914
QY 1052 ctcttacaanaattcctaagaagatttgaataaacaacttactattgttgcattactt 1111
DB 915 ctcttacaanaattcctaagaagatttgaataaacaacttactattgttgcattactt 974
QY 1112 gctctccatgatacactcgcagagttctctgagcgtctatcatcaacttattacggcac 1171
DB 975 gctctccatgatacacttgcagagttctctgagcgtctatcatcaacttattacggcac 1034
QY 1172 caagtataagagatttccacttgcgttggaacacttggttacagtgtgaaacaagcttgg 1231
DB 1035 caagtataagagatttccacttgcgttggaacacttggttacagtgtgaaacaagcttgg 1094
QY 1232 attactaagtttttcttctgtagtgcataagtggttgcataagcctctgcttaccagtag 1291
DB 1095 attactaagtttttcttctgtagtgcataagtggttgcataagcctctgcttaccagtag 1154
QY 1292 aaggtcagaagagatttcttctcacaatggtctatcagcaggttctatcgaataatga 1351
DB 1155 aaggtcagaagagatttcttctcacaatggtctatcagcaggttctatcgaataatga 1214
QY 1352 aaactcttggaatgagaagaagtttggagaattggaattatatacctccttggcataat 1411
DB 1215 aaactcttggaatgagaagaagtttggagaattggaattatatacctccttggcataat 1274
QY 1412 gagccttggttacttccctccctggcagcacttccatccctcagttgagcaattgttt 1471
DB 1275 gagccttggttacttccctccctggcagcacttccatccctcagttgagcaattgttt 1334
QY 1472 aaacttggaagaattcagtttattcagctcactgataatgtagtgcctgctcatalaag 1531
DB 1335 aaacttggaagaattcagtttattcagctcactgataatgtagtgcctgctcatalaag 1394
QY 1532 tacttccaagttttaaattatgtagtgaagaaagcgttcttgagaagaagtaactacagatt 1591
DB 1395 tacttccaagttttaaattatgtagtgaagaaagcgttcttgagaagaagtaactacagatt 1454
QY 1592 ttatacaccacaacttcttctgtctgttctgttgcctcctaattgtaattctgggttaa 1651
DB 1455 ttatacaccacaacttcttctgtctgttctgttgcctcctaattgtaattctgggttaa 1514
QY 1652 gatattttaaattccctcagatgataagcggaagcttaaaacgaatttaaaaaagcttgga 1711
DB 1515 gatattttaaattccctcagatgataagcggaagcttaaaacgaatttaaaaaagcttgga 1574
QY 1712 aagaagcccaatttctggaagaaggtatgtggaagaacaaatcccatatgtcccccggagag 1771
DB 1575 aagaagcccaatttctggaagaaggtatgtggaagaacaaatcccatatgtcccccggagag 1634
QY 1772 ggtcacagtaatgtagatgaataaaggtgtcacaagcgtgcataaagaattcctaactgac 1831
DB 1635 ggtcacagtaatgtagatgaataaaggtgtcacaagcgtgcataaagaattcctaactgac 1694
QY 1832 catattattatgactctcagttcagttacaaagtatgctcgaataattacgttgggttga 1891
DB 1695 catattattatgactctcagttcagttacaaagtatgctcgaataattacgttgggttga 1754

RESULT 2

Oy	1472	aaactgagagaaatcagtttttaacgtctacacttggatattgcgtctgcctcataag	1531
Db	1335	aaactggagagaattcagtttttaacgtctacacttggatattgcgtctgcctcataag	1394
Oy	1532	tacttccagttttaattatgatgagaaacagagtttttggagaagtgactacagtt	1591
Db	1395	tacttccagttttaattatgatgagaaacagagtttttggagaagtgactacagtt	1454
Oy	1592	ttaataccacccaactttctctgtctgttttgcctcaatttgaattcttggtaa	1651
Db	1455	ttataacacccaactttctctgtctgttttgcctcaatttgaattcttggtaa	1514
Oy	1652	gattatttattccttcctccatcgtataagccgaaagctaaacgaattaaanaagcttggaa	1711
Db	1515	gattatttattccttcctccatcgtataagccgaaagctaaacgaattaaanaagcttggaa	1574
Oy	1712	aaagaagccaattctctggaagaaggtatttggaggaacaattcctcatgtctcccggaag	1771
Db	1575	aaagaagccaattctctggaagaaggtatttggaggaacaattcctcatgtctcccggaag	1634
Oy	1772	ggtacagtaatgtagatataaagtgttccacagtcgcacataa	1817
Db	1635	ggtacagtaatgtagatataaagtgttccacagtcgcacataa	1680

RESULT	3	
AA515793	standard; DNA; 1725 BP.	
XX	AA515793	
XX	AA515793;	
XX	16-JAN-2002 (first entry)	
DT		
DE	Human DNA for Six-Transmembrane Protein of Prostate 1, STMP1.	
XX		
KW	Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;	
KW	benign prostatic hyperplasia; acute prostatitis; testicular cancer;	
KW	erythrocytosis; testicular disorder; proliferative disorder; lymphoma;	
KW	leukemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;	
KW	liver cancer; lung cancer; cyclostatic; ds.	
XX		
OS	Homo sapiens.	
XX		
XX	Key	Location/Qualifiers
FT	misc_feature	89
FT		/*tag= a
FT		/note= "Represents 338 nucleotides of intron 1"
FT	misc_feature	162
FT		/*tag= b
FT		/note= "Represents 12713 nucleotides of intron 2"
FT	CDS	200..1702
FT		/*tag= c
FT		/product= "STMP1"
FT	misc_feature	697
FT		/*tag= d
FT		/note= "Represents 1396 nucleotides of intron 3"
FT	misc_feature	1225
FT		/*tag= e
FT		/note= "Represents 2372 nucleotides of intron 4"
FT	misc_feature	1410
FT		/*tag= f
FT		/note= "Represents 2299 nucleotides of intron 5"
XX		
XX	MO200172962-A2.	
XX		
XX	04-OCT-2001.	
XX		
XX	23-MAR-2001; 2001WO-US09410.	
XX		
XX	24-MAR-2000; 2000US-191929P.	
XX		
XX	(SAAT/) SAATCIGLU F.	

XX Saac1cloglu F;
PI
XX
DR WPI: 2001-662926/76.
DR P-PSDB; AAU10187.
XX
PI New polynucleotide for the diagnosis, prevention and treatment for
PI prostate and testis disorders, particularly prostate cancer, comprises
PI prostate-specific or testis-specific nucleic acids
XX
PS Example 3; Fig 4B; 114pp; English.
XX
XX The invention relates to substantially pure prostate-specific or
CC testis-specific polypeptides and the nucleic acids encoding them.
CC Also included are vectors and host cells expressing the proteins, a
CC transgenic animal expressing the protein, antibodies against the
CC proteins, probes for detecting the nucleic acids, antisense molecules
CC for the nucleic acids and methods of isolating modulators of the
CC proteins. Compounds that modulate the prostate specific or testis
CC specific polypeptide are useful to diagnose, prevent or treat disorders
CC of the testis or prostate particularly prostate cancer, benign
CC prostatic hyperplasia, acute prostatitis, testicular cancer,
CC cryptorchidism, undescended, retractile, ascending or vanished
CC testis. Other proliferative disorders for which the modulators may be
CC used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
CC cancer, pancreatic cancer, liver cancer and lung cancer. The
CC present sequence encodes a prostate specific protein, Six-Transmembrane
CC Protein of Prostate 1, STMP1.
XX
XX Sequence 1725 BP; 476 A; 340 C; 387 G; 517 T; 5 other:
SQ

Query Match	77.5%	Score 1466.2	DB 22	Length 1725
Best Local Similarity	97.9%	Pred. No. 0		
Matches 1528	Conservative	0	Mismatches 3	Indels 30
				Gaps 3
Oy	280	aggaatctcttgatccttggaagtgccgcatcatgaatcctatgatggaa	339	
Db	165	aggaatccttggtatccttggaaatgcccgcatactgaaatcctatgatggaa	224	
Oy	340	gccctaagagcctagtgaacttgcttaacctatgcaaaatgltacaaagtgcga	399	
Db	225	gccctaagagcctagtgaacttgcttaacctatgcaaaatgltacaaagtgcga	284	
Oy	400	ggaagtgcactgtaagtgcattggaaatggagatcttgcgaatcctgacattcgac	459	
Db	285	ggaagtgcactgtaagtgcattggaaatggagatcttgcgaatcctgacattcgac	344	
Oy	460	ttatatagtcgcgcatacatgctgcgcataagagtagaaatcctaaattgcttcgaat	519	
Db	345	ttatagatgsgcatcatcatgtgcatatgagtagaaatcctaaattgcttcgaat	404	
Oy	520	ttttccctcaatgtagtagatgtaactcatcatgaagaatgctctcacaaaacaataa	579	
Db	405	ttttccctcaatgtagtagatgtaactcatcatgaagaatgctctcacaaaacaataa	464	
Oy	500	tattgttgctatacaacgagaacatttaactccctctggggaacctggaacctctgctg	639	
Db	465	tattgttgctatacaacgagaacatttaactccctctggggaacctggaacctctgctg	524	
Oy	640	tgggtaaatccctgattcatgtgagacaataaactgagataaacacgataccagaatcca	699	
Db	525	tgggtaaatccctgattcatgtgagacaataaactgagataaacacgataccagaatcca	584	
Oy	700	atgcgtcaatatttgcttcaatattccagattcttggattgltcacaagaattaaatgtg	759	
Db	585	atgcgtcaatatttgcttcaatattccagattcttggattgltcacaagaattaaatgtg	644	
Oy	760	tctcagcttgggcaactcaatgtagaactaaagatggccagccg-----caagtt	809	
Db	645	tctcagcttgggcaactcaatgtagaactaaagatggccagccgcaatgataagacagtt	704	
Oy	810	tataatcacagaacaatattcaacgcgcagacaacagatattgaaattccgcacatg	869	

```

|||||
Db 705 tataatagcagcaaatattcaagcgcgcagcaacaggttatgtacttgcgcgcaggttg 764
QY 870 aatttattccattgaacttggtatcccttatcatcagcagagagatgtgaaattacc 929
Db 765 aatttattccattgaacttggtatcccttatcatcagcagagagatgtgaaattacc 824
QY 930 ctacgactcttactctctcgtgagagagcaggttggttaacttaagcttgccacatt 989
Db 825 ctacgactcttactctctcgtgagagagcaggttggttaacttaagcttgccacatt 884
QY 990 ttttcttattctctctcgtgagagatgtgattcatccatgtcagaacacacagagt 1049
Db 885 ttttcttattctctctcgtgagagatgtgattcatccatgtcagaacacacagagt 944
QY 1050 gacttttaacaaattcctatatagagattgtgataaacccttactatgttgcattact 1109
Db 945 gacttttaacaaattcctatatagagattgtgataaacccttactatgttgcattact 1004
QY 1110 ttgctctccctagatatacctcgcaggtctctctgcagctgcttataacttattacgc 1169
Db 1005 ttgctctccctagatatacctcgcaggtctctctgcagctgcttataacttattacgc 1064
QY 1170 accaaglatagagagattccacacttggttgaaacctggttacaagtgtagaacaacagct 1229
Db 1065 accaaglatagagagattccacacttggttgaaacctggttacaagtgtagaacaacagct 1124
QY 1230 ggaattacaagttttctctcgtcatggtccatggttgcctacagcctctgcttaccgatg 1289
Db 1125 ggaattacaagttttctctcgtcatggtccatggttgcctacagcctctgcttaccgatg 1184
QY 1290 agaaagtcagaagagatattgtttccacaatgcttatacgc-----aggttca 1339
Db 1185 agaaagtcagaagagatattgtttccacaatgcttatacgc-----aggttca 1244
QY 1340 tgaagaatattgaaacactcttggaatagagaagaagatttggaagattgaaatgatatctc 1399
Db 1245 tgaagaatattgaaacactcttggaatagagaagaagatttggaagattgaaatgatatctc 1304
QY 1400 ctttgacataatgagccttggttacttccctcctgcgcagctactctatccctcaggt 1459
Db 1305 ctttgacataatgagccttggttacttccctcctgcgcagctactctatccctcaggt 1364
QY 1460 gagcaatgctttaacttgtagagaatcagtttatttcag-----tctaaccttg 1509
Db 1365 gagcaatgctttaacttgtagagaatcagtttatttcag-----tctaaccttg 1424
QY 1510 gataatgcgtcgtcgcataatgacttccatgttttaattatgatatgaaacagagctt 1569
Db 1425 gataatgcgtcgtcgcataatgacttccatgttttaattatgatatgaaacagagctt 1484
QY 1570 ttgaagagaagatctacacagatttatatacacacacaaacttgcttctgctctgtttgc 1629
Db 1485 ttgaagagaagatctacacagatttatatacacacacaaacttgcttctgctctgtttgc 1544
QY 1630 cctcaattgtaattcgttggaatagatttatttcttccatgatatagccggaactaa 1689
Db 1545 cctcaattgtaattcgttggaatagatttatttcttccatgatatagccggaactaa 1604
QY 1690 aacgaattaaanaaggctgggaaagaagcaatttctggaagaagttatggggagacaa 1749
Db 1605 aacgaattaaanaaggctgggaaagaagcaatttctggaagaagttatggggagacaa 1664
QY 1750 ttccctcatgtctcccgagagaggttcacagtaatgtgatataaatggtgttcacagctg 1809
Db 1665 ttccctcatgtctcccgagagaggttcacagtaatgtgatataaatggtgttcacagctg 1724
QY 1810 c 1810
Db 1725 c 1725

```

RESULT 4

```

AAD07072
ID AAD07072 standard; cDNA; 2453 BP.
XX AC AAD07072;
XX AC 06-AUG-2001 (first entry)
DE DT
XX DE Human six transmembrane epithelial antigen of prostate-2 clone GRD3 cDNA.
XX XX Human; cytosolic; antiproliferative; vaccine; gene therapy;
XX KW six transmembrane epithelial antigen of the prostate-2; STEAP-2;
XX KW chromosome 7q21; cancer; prostate; colon; bladder; lung; ovarian;
XX KW pancreatic; ss.
XX OS Homo sapiens.
XX XX
XX FH Key Location/Qualifiers
XX FH RBS 352..358
XX FT /tag= a
XX FT /note= "Kozak region"
XX FT CDS 355..1719
XX FT /tag= b
XX FT /product= "Human six transmembrane epithelial antigen
of the prostate (STEAP)-2"
XX FT CDS 709..2073
XX FT /tag= c
XX FT /product= "Human six transmembrane epithelial antigen
of the prostate (STEAP)-2, alternative version"
XX FT /note= "CDS does not include start and stop codon"
XX FT /transl_except= (pos:1714..1722, aa:Asp-Ala)
XX FT /transl_except= (pos:1834..1842, aa:Arg-Ser)
XX FT /transl_except= (pos:1957..1965, aa:Glu-Gly)
XX FT /transl_except= (pos:2050..2058, aa:Thr-Ser)
XX FT /transl_except= (pos:2062..2070, aa:Asn-Phe)
XX FT /note= "inframe stop codon alters the reading frame"
XX FT /partial
XX PN MO200140276-A2.
XX PD
XX PD 07-JUN-2001.
XX PF 06-DEC-2000; 2000MO-US33040.
XX PR 06-DEC-1999; 99US-0455486.
XX PA (UROC-) UROGENESYS INC.
XX PA Afar DEH, Hubert RS, Raitano AB, Saffran DC, Mitchell SC, Paris M;
PI Jakobovits A;
PI PI
XX DR WPI; 2001-367804/38.
XX DR P-PSDB; AAE02781, AAE02841.
XX PT New STEAP (six transmembrane epithelial antigen of the prostate)
XX PT proteins, expressed in human cancers, useful for detecting and treating
XX PT cancer.
XX PS Claim 4; Fig 9A-9D; 187pp; English.
XX PS
XX CC The present sequence is human six transmembrane epithelial antigen of
the prostate (STEAP)-2 clone GRD3 cDNA. STEAP is a member of cell
surface serpentine transmembrane antigens. STEAP-2 gene is located on
chromosome 7q21 and is used in gene therapy. Inhibiting the development
or progression of a cancer (eg. prostate, colon, bladder, lung, ovarian
and pancreatic) expressing STEAP or inhibiting growth or killing cells
expressing STEAP in a patient, comprises administering a vaccine
composition to the patient. Treating a patient with a cancer that
expresses STEAP, or inhibiting growth or killing cells expressing STEAP,
comprises administering to the patient a vector encoding single chain
monoclonal antibody that comprises the variable domains of the heavy and
light chains of the monoclonal antibody that specifically binds to STEAP,
such that the vector delivers the single chain monoclonal antibody coding
sequence to the cancer cells and the encoded single chain monoclonal

```

CC antibody is expressed intracellularly.
CC Note: The present sequence is also shown in sequence listing of the
specification, but it lacks nucleotides at its 5' end.
XX
SQ Sequence 2453 BP; 648 A; 537 C; 520 G; 748 T; 0 other;

Query Match	74.6%;	Score 1409.8;	DB 22;	Length 2453;
Best Local Similarity	93.0%;	Pred. No. 0;		
Matches 1534;	Conservative	0;	Mismatches 2;	Indels 113;
				Gaps 1.

QY	1	ggggaagcagctgtagagtgagcagccgcaacgcgagcagccacccgcaacccgcaacccgcaacgctgagagtg	60
Db	154	gggggaagcagctgtagagtgagcagccgcaacgcgagcagccacccgcaacccgcaacccgcaacgctgagagtg	213
QY	61	cagctccagcagcagccctcgagcccccgggtgtggccctctggggagctcgagcagcctcccgagag	120
Db	214	cagctccagcagcagccctcgagcccccgggtgtggccctctggggagctcgagcagcctcccgagag	273
QY	121	ctgcgaagctcgccccctcgccccgcgtggaagcgcgggggcgcgaggaagtgaagaga	180
Db	274	ctgcgaagcctcgccccctcgccccgcgtggaagcgcgggggcgcgaggaagtgaagaga	319
QY	181	ggaaaatggaaaatgtgagtgagccttcctgatactgctctcctctgctgtagaagggg	240
Db	320	-----	319
QY	241	aaagaaatcgatgatatattacagcgctccctataatacaaggaatctctgtgatactg	300
Db	320	-----	340
QY	301	gaagctccgctatataatgaaatcaatctcctatgagtggaagccctaaagagcttagtgaag	360
Db	341	gaagctccgctatataatgaaatcaatctcctatgagtggaagccctaaagagcttagtgaag	400
QY	361	ctgtcttaacctaagcgcaataatgtatcaaaaagaatgcaagaagagctacgtgatgtga	420
Db	401	ctgtcttaacctaagcgcaataatgtatcaaaaagaatgcaagaagagctacgtgatgtga	460
QY	421	ttagaagtgtgagatcttggccaatctctgaccatctgactatataatgctgcgctatacatg	480
Db	461	ttagaagtgtgagatcttggccaatctctgaccatctgactatataatgctgcgctatacatg	520
QY	481	tgtcataatggagaatgaataatcctaagttgtgcttctgataattttcccatctgtgtagatg	540
Db	521	tgtcataatggagaatgaataatcctaagttgtgcttctgataattttcccatctgtgtagatg	560
QY	541	tcaactcatcatgaagatgctctcccaaaaacaatatataatctgtgtcatcacagag	600
Db	581	tcaactcatcatgaagatgctctcccaaaaacaatatataatctgtgtcatcacagag	640
QY	601	aacattatactccctgtagggagccttgagacatctgctgtgtgggttaaatccctgattgag	660
Db	641	aacattatactccctgtagggagccttgagacatctgctgtgtgggttaaatccctgattgag	700
QY	661	tgaagcaataaataatgagataaacaacagatcaccaatgctgcgaatacttgctctcat	720
Db	701	tgaagcaataaataatgagataaacaacagatcaccaatgctgcgaatacttgctctcat	760
QY	721	tattccagatctcttgattgttcaaaagatttaagtgtgtccagctctggcaccttcagt	780
Db	761	tattccagatctcttgattgttcaaaagatttaagtgtgtccagctctggcaccttcagt	820
QY	781	taggaacctaagatgagccaagccggaggttatatatatgcaagcaacaatatccaagccgcagc	840
Db	821	taggaacctaagatgagccaagccggaggttatatatatgcaagcaacaatatccaagccgcagc	880
QY	841	aacaggttatgaaactgtcccgccagcttggaattcaatcccatgtgattggatctctat	900
Db	881	aacaggttatgaaactgtcccgccagcttggaattcaatcccatgtgattggatctctat	940
QY	901	catcagccagagagatgtaaaattacccctcagactcttactctctctgtagagggccag	960

Db	941	catcaagccagagagatgtgaaatttacccttaacgacttctaactctctcgagagagccag	1000
Qy	961	tgtgtgtgatacctaagcttggccacattttttctctatctcccttggtcagagatgta	1020
Db	1001	tgtgtgtgatacctaagcttggccacattttttctctatctcccttggtcagagatgta	1060
Qy	1021	ttcatcatalgtctagaacaaccaaagatgacttttaacaaaattccatagagattgtga	1080
Db	1061	ttcatcatalgtctagaacaaccaaagatgacttttaacaaaattccatagagattgtga	1120
Qy	1081	ataaaaccttaacctaatagttgcacatttctctccctgatacctcgagagcttc	1140
Db	1121	ataaaaccttaacctaatagttgcacatttctctccctgatacctcgagagcttc	1180
Qy	1141	tggagactgtctatccaacttattacgagccaaatatalagagatttccacctgtgttg	1200
Db	1181	tggagactgtctatccaacttattacgagccaaatatalagagatttccacctgtgttg	1240
Qy	1201	aaacctgtgttaacagttgtagaaaaacagcttggattactaatgtttttctctgctatgtcc	1260
Db	1241	aaacctgtgttaacagttgtagaaaaacagcttggattactaatgtttttctctgctatgtcc	1300
Qy	1261	atgtgtgctcaagcctctgtgttaccgatagagaaggtcagagagatattgtttctcaaca	1320
Db	1301	atgtgtgctcaagcctctgtgttaccgatagagaaggtcagagagatattgtttctcaaca	1360
Qy	1321	tgtgttalcagcaggtgtcalcagcaaatatgtgaaaactcttggaaatgagagaagaattgtga	1380
Db	1361	tgtgttalcagcaggtgtcalcagcaaatatgtgaaaactcttggaaatgagagaagaattgtga	1420
Qy	1381	gaattgaaatgtatattctcttggcattatgaagccttggcttacttccctctgtgcag	1440
Db	1421	gaattgaaatgtatattctcttggcattatgaagccttggcttacttccctctgtgcag	1480
Qy	1441	tcaacttaccctctgaagtgagaaatgcttttaactctggagagaatcagtttaattcagt	1500
Db	1481	tcaacttaccctctgaagtgagaaatgcttttaactctggagagaatcagtttaattcagt	1540
Qy	1501	ctaaccttggataltgtctgtctgtctatagaacttccacagtttctaattatagatgta	1560
Db	1541	ctaaccttggataltgtctgtctgtctatagaacttccacagtttctaattatagatgta	1600
Qy	1561	aacgagcttttggagaagagtaactaagaatttatataccaccaaacttgttctgtctc	1620
Db	1601	aacgagcttttggagaagagtaactaagaatttatataccaccaaacttgttctgtctc	1660
Qy	1621	ttgtttgcctcaattgtaattctgtgt	1649
Db	1661	ttgtttgcctcaattgtaattctgtgt	1689
RESULT 5			
AAS15811			
ID	AAS15811 standard; cDNA; 1561 BP.		
AC	AAS15811;		
XX	16-JAN-2002 (first entry)		
DT			
XX	Human cDNA encoding ORF2 of Six-Transmembrane Protein of Prostate 1.		
DE			
XX			
KW	Human; six-Transmembrane Protein of prostate 1; STMP1; prostate cancer		
KW	benign prostatic hyperplasia; acute prostatitis; testicular cancer;		
KW	cryptorchidism; testicular disorder; proliferative disorder; lymphoma;		
KW	leukemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;		
KW	liver cancer; lung cancer; cytosolic; ss; ORF2.		
XX			
OS	Homo sapiens.		
XX			
EH	Key		
FT	Location/Qualifiers		
CDS	188..1552		
FT	/tag= a		
FT	/product= "STMP1 ORF2"		

XX MO200172962-A2.
XX 04-OCT-2001.
XX 23-MAR-2001; 2001MO-US09410.
XX 24-MAR-2000; 2000US-191929P.
XX (SAAT/) SAATC10GLU F.
XX Saatl0glu F;
XX MPI; 2001-662926/76.
XX P-PSDB; AAU10188.
XX New polynucleotide for the diagnosis, prevention and treatment for
XX prostate and testis disorders, particularly prostate cancer, comprises
XX prostate-specific or testis-specific nucleic acids
XX
XX Claim 4; Fig 4H; 11App; English.
XX
XX The invention relates to substantially pure prostate-specific or
XX testis-specific polypeptides and the nucleic acids encoding them.
XX Also included are vectors and host cells expressing the proteins, a
XX transgenic animal expressing the protein, antibodies against the
XX proteins, probes for detecting the nucleic acids, antisense molecules
XX for the nucleic acids and methods of isolating modulators of the
XX proteins. Compounds that modulate the prostate specific or testis
XX specific polypeptide are useful to diagnose, prevent or treat disorders
XX of the testis or prostate particularly prostate cancer, benign
XX prostatic hyperplasia, acute prostatitis, testicular cancer,
XX cryptorchidism, undescended, retractile, ascending or vanished
XX testis. Other proliferative disorders for which the modulators may be
XX used include lymphoma, leukemia, melanoma, ovarian cancer, breast
XX cancer, pancreatic cancer, liver cancer and lung cancer. The
XX present sequence encodes prostate specific protein, Six-Transmembrane
XX Protein of Prostate 1, STMP1, ORF2.
XX
XX Sequence 1561 BP; 420 A; 319 C; 346 G; 476 T; 0 other;
SQ
Query Match 72.4%; Score 1369.2; DB 22; Length 1561;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1371; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 276 tcaagagatattcttgtagcttggaagtgctcgatcatcagtaaatcctatgatg 335
DB 149 tccagagatatcttgtagcttggaagtgctcgatcatcagtaaatcctatgatg 208
QY 336 ggaagccctaagagccttgtagaacttggttaccctaagtgcataatgatacaaat 395
DB 209 ggaagccctaagagccttgtagaacttggttaccctaagtgcataatgatacaaat 268
QY 396 gcaaggaagtcacgttagtgatgtgaagtggaagattgccaaatcctggaccat 455
DB 269 gcaaggaagtcacgttagtgatgtgaagtggaagattgccaaatcctggaccat 328
QY 456 cgactattagatgcgcatcatcagtgcataagagtagaagtaaacctaaagttgcttc 515
DB 329 cgactattagatgcgcatcatcagtgcataagagtagaagtaaacctaaagttgcttc 388
QY 516 gaatttttctcatgtgtgaagtgcacatcatcagaagatgctctcaaaaaacaat 575
DB 389 gaatttttctcatgtgtgaagtgcacatcatcagaagatgctctcaaaaaacaat 448
QY 576 ataataattgtctatcacagaagaacattatacctccctgaggagacctgagacatcg 635
DB 449 ataataattgtctatcacagaagaacattatacctccctgaggagacctgagacatcg 508
QY 636 ctgtgtggttaaaactcgtatgatgtgagcaataacatgagataaacagatccagaa 695
DB 509 ctgtgtggttaaaactcgtatgatgtgagcaataacatgagataaacagatccagaa 568

QY 696 tccaatgctgaatatattggtctcatattccagatctcttgattgtgcaaaagattaat 755
DB 569 tccaatgctgaatatattggtctcatattccagatctcttgattgtgcaaaagattaat 628
QY 756 gttgtcgaagcttgggcaacttcagtaggaacctaaagatgcagcgaggtttatata 815
DB 629 gttgtcgaagcttgggcaacttcagtaggaacctaaagatgcagcgaggtttatata 688
QY 816 tgcagcaacaatatattcaagcgagcaacaagattatgtaacttgcgcgagttgaattc 875
DB 689 tgcagcaacaatatattcaagcgagcaacaagattatgtaacttgcgcgagttgaattc 748
QY 876 attccattgacttggagaccttatcatcagccagagaagattgaaatttaacctcaga 935
DB 749 attccattgacttggagaccttatcatcagccagagaagattgaaatttaacctcaga 808
QY 936 ctcttactctctggaagagggcagtggtgtagtataagcttggccaatctttttc 995
DB 809 ctcttactctctggaagagggcagtggtgtagtataagcttggccaatctttttc 868
QY 996 cttaattccttgcagagatgtgataccataatgctagaacaacagagtgactt 1055
DB 869 cttaattccttgcagagatgtgataccataatgctagaacaacagagtgactt 928
QY 1056 tacaagaattcctatagagatgtggaataaaccttactactagttgcatcttctc 1115
DB 929 tacaagaattcctatagagatgtggaataaaccttactactagttgcatcttctc 988
QY 1116 tccctagtaactctcgcaagctctctctgcaagctgctataacttatacgcacaag 1175
DB 989 tccctagtaactctcgcaagctctctctgcaagctgctataacttatacgcacaag 1048
QY 1176 tataagagattccacctgtgtggaacctggtttacagtgtagaagaaacagcttgatta 1235
DB 1049 tataagagattccacctgtgtggaacctggtttacagtgtagaagaaacagcttgatta 1108
QY 1236 ctgaatttttctctgctatgctgacatggttgcataagcctctgcttaccagtagaagg 1295
DB 1109 ctgaatttttctctgctatgctgacatggttgcataagcctctgcttaccagtagaagg 1168
QY 1296 tcagagagatatattgttctccaacatggtctatcagcaggttcatgcaaatatgaaac 1355
DB 1169 tcagagagatatattgttctccaacatggtctatcagcaggttcatgcaaatatgaaac 1228
QY 1356 tcttggaaatggaagaagtttggagaattgaaatgataatcctcttggcataatgagc 1415
DB 1229 tcttggaaatggaagaagtttggagaattgaaatgataatcctcttggcataatgagc 1288
QY 1416 ctggcttacttccctcccgagcagatccttataccttcaatgtagcaattgctttaaac 1475
DB 1289 ctggcttacttccctcccgagcagatccttataccttcaatgtagcaattgctttaaac 1348
QY 1476 tggagagaattcagtttattatcagctcacaactggaatgctgcctgcgcataagtaact 1535
DB 1349 tggagagaattcagtttattatcagctcacaactggaatgctgcctgcgcataagtaact 1408
QY 1536 ttccatgttttaattatgataatgaaacagagcttttgaaggaagatgactacagatttcat 1595
DB 1409 ttccatgttttaattatgataatgaaacagagcttttgaaggaagatgactacagatttcat 1468
QY 1596 acacacacaaacttgttctgtctgttgccttggtttgcctcaatttgaatttcgggt 1649
DB 1469 acacacacaaacttgttctgtctgttgccttggtttgcctcaatttgaatttcgggt 1522
RESULT 6
AA515810
ID AA515810 standard; cDNA; 2238 BP.
XX
AC AA515810;
XX
DT 16-JAN-2002 (first entry)

|||||
Db 1409 ttccatgtttaattatgatgagaaagacgttttgaggaagagactacacagatttat 1468
QY 1596 aaaccccaacttcttctgtctctgtttgcccataatgtaattcttgat 1649
|||||
Db 1469 aaaccaccaacttcttctgtctctgtttgcccataatgtaattcttgat 1522

RESULT 7

AA515813
ID AA515813 standard; cDNA; 2102 BP.

AA515813;

16-JAN-2002 (first entry)

Human cDNA encoding ORF3 of Six-Transmembrane Protein of Prostate 1.

XX Human; Six-Transmembrane Protein of Prostate 1; STMP1; prostate cancer;
KW benign prostatic hyperplasia; acute prostatitis; testicular cancer;
KW cryptorchidism; testicular disorder; proliferative disorder; lymphoma;
KW leukaemia; melanoma; ovarian cancer; breast cancer; pancreatic cancer;
KW liver cancer; lung cancer; cytostatic; SS; ORF3.

XX Homo sapiens.

XX Key location/Qualifiers

XX FT CDS 171..1430

XX FT /*tag= a

XX FT /product= "STMP1, ORF3"

XX PN WO200172962-A2.

XX PD 04-OCT-2001.

XX PF 23-MAR-2001; 2001WO-US09410.

XX PR 24-MAR-2000; 2000US-191929P.

XX PA (SAAT/) SAATCTOGLU F.

XX PI Saactioglu F;

XX DR WPI: 2001-662926/76.

XX DR P-PSDB: ANU10189.

XX PT New polynucleotide for the diagnosis, prevention and treatment for
XX prostate and testis disorders, particularly prostate cancer, comprises
XX prostate-specific or testis-specific nucleic acids

XX PS Claim 4; Fig 4K; 114pp; English.

XX The invention relates to substantially pure prostate-specific or
XX testis-specific polypeptides and the nucleic acids encoding them.
XX Also included are vectors and host cells expressing the proteins, a
XX transgenic animal expressing the protein, antibodies against the
XX proteins, probes for detecting the nucleic acids, antisense molecules
XX for the nucleic acids and methods of isolating modulators of the
XX proteins. Compounds that modulate the prostate specific or testis
XX specific polypeptide are useful to diagnose, prevent or treat disorders
XX of the testis or prostate particularly prostate cancer, benign
XX prostatic hyperplasia, acute prostatitis, testicular cancer,
XX cryptorchidism, undescended, retractile, ascending or vanished
XX testis. Other proliferative disorders for which the modulators may be
XX used include lymphoma, leukaemia, melanoma, ovarian cancer, breast
XX cancer, pancreatic cancer, liver cancer and lung cancer. The
XX present sequence encodes prostate specific protein, Six-Transmembrane
XX Protein of Prostate 1, STMP1, ORF3.

XX Sequence 2102 BP; 595 A; 424 C; 422 G; 661 T; 0 other;

Query Match

64.3%; Score 1216.8; DB 22; Length 2102;

Best Local Similarity 99.4%; Pred. No. 0;
Matches 1221; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 272 atattcaagagatattcttgggtatctcttggaagtgctccgatatcatgtaatcattat 331
|||
Db 128 atgtctcagagatattcttgggtatctcttggaagtgctccgatatcatgtaatcattat 187
QY 332 gatgggaagccctaaagagccttagtgaactctgttaccatgaagcctaataatgatac 391
|||||
Db 188 gatgggaagccctaaagagccttagtgaactctgttaccatgaagcctaataatgatac 247
QY 392 agatgcaaggaagtgctacgtgaagtgtgattggaagtggaagatttgccaatccctgac 451
|||||
Db 248 agatgcaaggaagtgctacgtgaagtgtgattggaagtggaagatttgccaatccctgac 307
QY 452 catcgactatattagatgcgggtatacagtgtgtcacaaggaagataatccctaagttgc 511
|||||
Db 308 catcgactatattagatgcgggtatacagtgtgtcacaaggaagataatccctaagttgc 367
QY 512 ttctgaatttttcccatggtgtgattgtcatctatcatgaagatgcctcccaaaaac 571
|||||
Db 368 ttctgaatttttcccatggtgtgattgtcatctatcatgaagatgcctcccaaaaac 427
QY 572 aaataataattgtgtctatacacagagaacacataaacctccgttggaacctgagaca 631
|||||
Db 428 aaataataattgtgtctatacacagagaacacataaacctccgttggaacctgagaca 487
QY 632 tctgtcttgggttaaaatccctgatgtgattggaacataacatgagataaaccagtaacc 691
|||||
Db 488 tctgtcttgggttaaaatccctgatgtgattggaacataacatgagataaaccagtaacc 547
QY 692 agaattccaatgtcgaatattggtcctcatattcccaagattcttgattgtcaaaagatt 751
|||||
Db 548 agaattccaatgtcgaatattggtcctcatattcccaagattcttgattgtcaaaagatt 607
QY 752 taatgtgtcgaagcttgggcaacttcaagttaggaacctgaagatgccaagcggaggttta 811
|||||
Db 608 taatgtgtcgaagcttgggcaacttcaagttaggaacctgaagatgccaagcggaggttta 667
QY 812 tataatgagacaacaattccaagcgcgacaacagattatgaacttgcgcgcagttgaa 871
|||||
Db 668 tataatgagacaacaattccaagcgcgacaacagattatgaacttgcgcgcagttgaa 727
QY 872 ttcatctccatgacttggatggatcccttatcatcagccagaagatgaaatttaccct 931
|||||
Db 728 ttcatctccatgacttggatggatcccttatcatcagccagaagatgaaatttaccct 787
QY 932 acgaacttactctctgagagagggcgagtggtgtgagatatgaagcttggccacatttt 991
|||||
Db 788 acgaacttactctctgagagagggcgagtggtgtgagatatgaagcttggccacatttt 847
QY 992 ttctcttatctccttgcagaagatgtgattcatcatatgctagaacaacaagatga 1051
|||||
Db 848 ttctcttatctccttgcagaagatgtgattcatcatatgctagaacaacaagatga 907
QY 1052 ctttacaatatccctatagagattgtgaataaaccctaccatagttgccaacttt 1111
|||||
Db 908 ctttacaatatccctatagagattgtgaataaaccctaccatagttgccaacttt 967
QY 1112 gctctccctagatatccctgcaggtctcttgcgaagcttcatcaacttatatagcgac 1171
|||||
Db 968 gctctccctagatatccctgcaggtctcttgcgaagcttcatcaacttatatagcgac 1027
QY 1172 caagtatagagattccaaccttgggtggaacactgtgtacagtgtaagaanaaagcttgg 1231
|||||
Db 1028 caagtatagagattccaaccttgggtggaacactgtgtacagtgtaagaanaaagcttgg 1087
QY 1232 attactaagtttttcttgcatagtgtcattgttgccaagcctctgtcttaccagttgag 1291
|||||
Db 1088 attactaagtttttcttgcatagtgtcattgttgccaagcctctgtcttaccagttgag 1147
QY 1292 aaggtcagaagatatgtttctcacaatggttatcagcaggtcatgcaaatattga 1351
|||||

Db	1148	aaggtcagagagatattgtttctcaacatgacctacgcagggttcaatgcaaatltga	1207
Qy	1352	aaactcttgaagagagaagtttggagaattgaattatatactcctttggcataat	1411
Db	1208	aaactcttgaatgagagaagtttggagaattgaattatatactcctttggcataat	1267
Qy	1412	gagccttggcttaacttccctcctgcagcaactctataccttcgaatggcaatgctt	1471
Db	1268	gagccttggcttacttccctcctgcagcaactctataccttcgltgagcaatgctt	1327
Qy	1472	aaacttgagagaattcaattttatcaag	1499
Db	1328	aaacttgagagaattcaattttatcaag	1355
RESULT 8			
AAS76493			
ID	AAS76493 standard; cDNA; 2192 BP.		
XX			
AC	AAS76493;		
XX			
DT	13-FEB-2002 (first entry)		
XX			
DE	DNA encoding novel human diagnostic protein #12297.		
XX			
KW	Human; Chromosome mapping; gene mapping; gene therapy; forensic;		
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200175067-A2.		
PD	11-OCT-2001.		
XX			
PF	30-MAR-2001; 2001MO-US08631.		
XX			
PR	31-MAR-2000; 2000US-0540217.		
PR	23-AUG-2000; 2000US-0649167.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
P1	Dmanac RT, Liu C, Tang YT;		
XX			
DR	WPI; 2001-639362/73.		
XX			
PT	P-PSDB; ABG12306.		
XX			
PT	New isolated polynucleotide and encoded polypeptides, useful in		
PT	diagnostics, forensics, gene mapping, identification of mutations		
PT	responsible for genetic disorders or other traits and to assess		
PT	biodiversity -		
PS	Claim 1; SEQ ID No 12297; 103pp; English.		
XX			
CC-	The invention relates to isolated polynucleotide (I) and		
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,		
CC	(polymerase chain reaction (PCR) primers, oligomers, and for chromosome		
CC	and gene mapping, and in recombinant production of (II). The		
CC	polynucleotides are also used in diagnostics as expressed sequence tags		
CC	for identifying expressed genes. (I) is useful in gene therapy techniques		
CC	to restore normal activity of (II) or to treat disease states involving		
CC	(II). (II) is useful for generating antibodies against it, detecting or		
CC	quantitating a polypeptide in tissue, as molecular weight markers and as		
CC	a food supplement. (II) and its binding partners are useful in medical		
CC	imaging of sites expressing (II). (I) and (II) are useful for treating		
CC	disorders involving aberrant protein expression or biological activity.		
CC	The polypeptide and polynucleotide sequences have applications in		
CC	diagnostics, forensics, gene mapping, identification of mutations		
CC	responsible for genetic disorders or other traits to assess biodiversity		
CC	and to produce other types of data and products dependent on DNA and		
CC	amino acid sequences. AAS64197-AAS94564 represent novel human		
CC	diagnostic coding sequences of the invention.		
CC	Note: The sequence data for this patent did not appear in the printed		
CC	specification, but was obtained in electronic format directly from WIPO		

QY	961	tggtggtgagctataagcttggccacatttttttccctattcccttgcagagatgtga	1020
Db	941	tgtgtgtgagctataaagcttggccacatttttttccctattcccttgcagagatgtga	1000
QY	1021	ttcatccatatgctcagaanaaccacagagacttttaacaaattccctataagatgtga	1080
Db	1001	ttcatccatatgctcagaanaaccacagagacttttaacaaattccctataagatgtga	1060
QY	1081	ataaaaccttaccataagatgtgccattaccttgcctcccaatatacactcgcaagcttgc	1140
Db	1061	ataaaaccttaccataagatgtgccattaccttgcctcccaatatacactcgcaagcttgc	1120
QY	1141	tggtgagctgtcttatcaacttatataaggacccaagatagaggagattccaccttggatgg	1200
Db	1121	tgtgagctgtcttatcaacttatataaggacccaagatagaggagattccaccttggatgg	1180
QY	1201	aaacctgtgttacaggtgtagaanaacagcttggatattactaagttttttcttcgtctatggtcc	1260
Db	1181	aaacctgtgttacaggtgtagaanaacagcttggatattactaagttttttcttcgtctatggtcc	1240
QY	1261	atgttgcctacagctctcgttccgaatagaagaagtcagagaagatattgttttccaca	1320
Db	1241	atgttgcctacagctctcgttccgaatagaagaagtcagagaagatattgttttccaca	1300
QY	1321	tggtctatcagcag	1334
Db	1301	tgtgttatacagcag	1314

	RESULT	9
AAS64300	ID	AAS64300 standard; cDNA; 3900 BP.
XX AC AAS64300;		
XX DT 13-FEB-2002 (first entry)		
DE DNA encoding novel human diagnostic protein #104.		
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;		
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.		
OS Homo sapiens.		
PN WO200175067-A2.		
PD 11-OCT-2001.		
PF 30-MAR-2001; 2001WO-US08631.		
PR 31-MAR-2000; 2000US-0540217.		
PPR 23-AUG-2000; 2000US-0649167.		
PA (HYSE-) HYSEQ INC.		
XX Drmanac RT, Liu C, Tang YT;		
XPI WPT; 2001-639362/73.		
DRI P-PSTDB; ABG00113.		
DR New isolated polynucleotide and encoded peptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -		
PT Claim 1; SEQ ID No 104; 103bp; English.		
PS The invention relates to isolated polynucleotide (I) and XX polypeptide (II) sequences. (I) is useful as hybridisation probes, CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome CC and gene mapping, and in recombinant production of (II). The CC polynucleotides are also used in diagnostics as expressed sequence tags CC for identifying expressed genes. (I) is useful in gene therapy techniques		

to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 3900 BP; 1161 A; 798 C; 892 G; 1049 T; 0 other;

Query Match	57.8%	Score 1093.2	DB 23	Length 3900
Best Local Similarity	91.3%	Fred. No. 4.1e-314		
Matches 1218	Conservative 0	Mismatches 3	Indels 113	Gaps 14
Oy	1	ggggaagcagctctggagctgacccgacagccagcccttgcaacccgacgtccgaggtg	60	
Db	94	gggggaagcagctctggagctgacccgacagccagcccttgcaacccgacgtccgaggtg	153	
Oy	61	cagtcctgaagccctctgcccggctggggctctgggagctgcgcgcgtcccgagagag	120	
Db	154	cagtcctgaagccctctgcccggctggggctctgggagctgcgcgcgtcccgagagag	213	
Oy	121	ctgcgaagctctgcccctgcggcgcttgagggcgcgggggcgcgagagaatgaagaga	180	
Db	214	ctgcgaagctctgcccctgcggcgcttgagggcgcgggggcgcgagagaatgaagaga	259	
Oy	181	ggaatttggaattgtgagtgagctcttgtaactgcctccctcgtgtgaaaaagggt	240	
Db	260	-----	259	
Oy	241	aaagaacgtcatgtatatatcagcgtcctctaatcaagaagatatctctgtgactgtg	300	
Db	260	-----	280	
Oy	301	gaagtgctccgtatcatgtgaatcattctctatgagtgggaagccctgaagccttaaggaa	360	
Db	281	gaagtgctccgtatcatgtgaatcattctctatgagtgggaagccctgaagccttaaggaa	340	
Oy	361	ctgtgttccttaatggcctaaatggatcatcaagaatggaaggttcactgtgagtgtga	420	
Db	341	cttttcttccttaatggcctaaatggatcatcaagaatggaaggttcactgtgagtgtga	400	
Oy	421	tggagagtggaagattctgcacaatctctgacacatcgcagcttatcttgatgcgtatcatg	480	
Db	401	tggagagtggaagattctgcacaatctctgacacatcgcagcttatcttgatgcgtatcatg	460	
Oy	481	tgtgtcataggaagtgaataatcctaagctgtgtcttcgtgaatttttccctaatgtgtgtagatg	540	
Db	461	tgtgtcataggaagtgaataatcctaagctgtgtcttcgtgaatttttccctaatgtgtgtagatg	520	
Oy	541	tcactcatcatgagaagtgtctctcacaaaaacaataataatatctgtgtataacacagag	600	
Db	521	tcactcatcatgagaagtgtctctcacaaaaacaataataatatctgtgtataacacagag	580	
Oy	601	aacattataccctccctgtgggacccgtgagacatctgctctgtgggtaaaatcccgattgag	660	
Db	581	aacattataccctccctgtgggacccgtgagacatctgctctgtgggtaaaatcccgattgag	640	
Oy	661	tgaagcaataacatgaggaataaacccgtatcccaagaatcccaatgtctgaatatctgtctcat	720	
Db	641	tgaagcaataacatgaggaataaacccgtatcccaagaatcccaatgtctgaatatctgtctcat	700	
Oy	721	tattcccaagattctttgattgtcaaaagatttaattgtgtgtcctcagcttgggcattcagct	780	

Db 701 taatccagatcttctgattgtcaagaatttaattgtgtctcagcttggcacttcagt 760
Qy 781 tagaaccataagatgcagccgagcttataatataatgacgaacataatcaagcgac 840
Db 761 tagaaccataagatgcagccgagcttataatataatgacgaacataatcaagcgac 820
Qy 841 aacaggtattgaacttgcgcgcagcttgaattcattccacttgaacttggatccctat 900
Db 821 aacaggtattgaacttgcgcgcagcttgaattcattccacttgaacttggatccctat 880
Qy 901 catcagccagagagatgtaaaattacccctacgacttctactctcggagagcgac 960
Db 881 catcagccagagagatgtaaaattacccctacgacttctactctcggagagcgac 940
Qy 961 tgggtgtagcataagcttgcgcacatttttccctattccttcttgcagagatgta 1020
Db 941 tgggtgtagcataagcttgcgcacatttttccctattccttcttgcagagatgta 1000
Qy 1021 ttcatcatatgtagaacaacacagagtgacttaccataatccctatagagattgta 1080
Db 1001 ttcatcatatgtagaacaacacagagtgacttaccataatccctatagagattgta 1060
Qy 1081 ataaacctaccataatgttcacattacttgcctcctcctagtagtaccctgcaggtctc 1140
Db 1061 ataaacctaccataatgttcacattacttgcctcctcctagtagtaccctgcaggtctc 1120
Qy 1141 tggcagctgtctatcaacttattacacgacacagatagagatttccacttggctg 1200
Db 1121 tggcagctgtctatcaacttattacacgacacagatagagatttccacttggctg 1180
Qy 1201 aaacctgttcaagtgtgtaaaacacagcttggattactaagtttttctcgtatgtcc 1260
Db 1181 aaacctgttcaagtgtgtaaaacacagcttggattactaagtttttctcgtatgtcc 1240
Qy 1261 atgttgcctacagcctcgtctacagatgagaagtgacagagatattgtttccaca 1320
Db 1241 atgttgcctacagcctcgtctacagatgagaagtgacagagatattgtttccaca 1300
Qy 1321 tggcttaccagcag 1334
Db 1301 tggcttaccagcag 1314

RESULT 10
ABA45820
ID ABA45820 standard; DNA; 1953 BP.
XX AC ABA45820;
XX DT 01-FEB-2002 (first entry)
XX DE Human breast cell single exon nucleic acid probe #4515.
XX KM Human; microarray; single exon probe; gene expression; breast;
XX KW disease; cancer; ss.
XX OS Homo sapiens.
XX PN WO200157271-A2.
XX PD 09-AUG-2001.
XX PE 30-JAN-2001; 2001WO-US00662.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI: 2001-496933/54.
XX
XX
XX New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes
XX
XX
PS Claim 1; SEQ ID NO 4515; 327pp + sequence listing; English.
XX
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at http://wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1953 BP; 633 A; 316 C; 324 G; 680 T; 0 other;
SQ

Query Match .28.1%; Score 530.6; DB 22; Length 1953;
Best Local Similarity 99.3%; Pred. No. 6,5e-147;
Matches 533; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 280 aggatattctgtgacttctggaagtgcgtatcatggaatcaatctatgtaggaa 339
Db 343 aggatattctgtgacttctggaagtgcgtatcatggaatcaatctatgtaggaa 402
Qy 340 gccctaaagagcttaagtaactgttttactaataatgacataatgataaagatgcaa 399
Db 403 gccctaaagagcttaagtaactgttttactaataatgacataatgataaagatgcaa 462
Qy 400 ggaaggtcacctgtgctgattggaagtggagatttgcgaatcccttgaaccattgac 459
Db 463 ggaaggtcacctgtgctgattggaagtggagatttgcgaatcccttgaaccattgac 522
Qy 460 ttattagatgcgctatcatgtgtgctacataggaagtagaataccttaagtcttctgaat 519
Db 523 ttattagatgcgctatcatgtgtgctacataggaagtagaataccttaagtcttctgaat 582
Qy 520 ttttctctatgtgtgattgctatcatatggaagtgtcttccaaaacaaataataa 579
Db 583 ttttctctatgtgtgattgctatcatatggaagtgtcttccaaaacaaataataa 642
Qy 580 tattgtgtctatacacagaaacataatacctcctgttggaccttggacatctgtctg 639
Db 643 tattgtgtctatacacagaaacataatacctcctgttggaccttggacatctgtctg 702
Qy 640 tgggtaaatcctgtatgtatgtgacataacataggaataaaccagttaccagaatcca 699
Db 703 tgggtaaatcctgtatgtatgtgacataacataggaataaaccagttaccagaatcca 762
Qy 700 atgttgataatttggcttcaatttcccaatcttggatttgcgaaggttaattgtg 759
Db 763 atgttgataatttggcttcaatttcccaatcttggatttgcgaaggttaattgtg 822
Qy 760 tctcagcttggcacttcaattagacataagatgacacgagcttataat 816
Db 823 tctcagcttggcacttcaattagacataagatgacacgagcttataat 879

```
RESULT 11
ABA56335
ID ABA56335 standard; DNA; 1953 BP.
XX
XX ABA56335;
XX
XX 01-FEB-2002 (first entry)
XX
XX Human foetal liver single exon nucleic acid probe #4640.
DE
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX Homo sapiens.
OS
XX MO200157277-A2.
PN
XX 09-AUG-2001.
PD
XX 30-JAN-2001; 2001WO-US00669.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT
XX analyzing gene expression in human foetal liver -
XX
XX Claim 1: SEQ ID NO 4640; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
CC
XX measuring human gene expression in a sample derived from human foetal
CC
XX liver. The single exon nucleic acid probes may be used for predicting,
CC
XX measuring and displaying gene expression in samples derived from human
CC
XX foetal liver. The present sequence is a single exon nucleic acid
CC
XX probe of the invention.
CC
XX Note: The sequence data for this patent did not form part of the
CC
XX printed specification, but was obtained in electronic format directly
CC
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1953 BP; 633 A; 316 C; 324 G; 680 T; 0 other;
SO

Query Match 28.1%; Score 530.6; DB 22; Length 1953;
Best Local Similarity 99.3%; Pred. No. 6.5e-147;
Matches 533; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 520 ttcttcctatgtgtgtagatgltcatcctcatcgaagaatgtctctcaaaaaaataataa 579
DB 583 ttcttcctatgtgtgtagatgltcatcctcatcgaagaatgtctctcaaaaaaataataa 642
QY 580 tatttgtgtatacaacagaagaacattatataccctccctgtgtggacctggaacattctgttg 639
DB 643 tatttgtgtatacaacagaagaacattatataccctccctgtgtggacctggaacattctgttg 702
QY 640 tgggtataaactcgtatgtagtgaagcaatacagtagaagtaaacacgctaccagaatcca 699
DB 703 tgggtataaactcgtatgtagtgaagcaatacagtagaagtaaacacgctaccagaatcca 762
QY 700 atgtgtaataattgtgtctcatattccacagattccttattgtgtgtcaaaaggaattaatgtg 759
DB 763 atgtgtaataattgtgtctcatattccacagattccttattgtgtgtcaaaaggaattaatgtg 822
QY 760 tctcagcttgggacacttcaagtagacctaagaatgccagcgcgaggttatata 816
DB 823 tctcagcttgggacacttcaagtagacctaagaatgccagcgcgaggttatatt 879

RESULT 12
ABA25976
ID ABA25976 standard; DNA; 1953 BP.
XX
XX ABA25976;
XX
XX 23-JAN-2002 (first entry)
XX
XX Probe #4442 for gene expression analysis in human heart cell sample.
DE
XX
XX Human; gene expression; heart; microarray; vascular system; probe;
KW
XX cardiovascular disease; hypertension; cardiac arrhythmia;
XX
XX congenital heart disease; ss.
XX
XX Homo sapiens.
OS
XX
XX MO200157274-A2.
PN
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00666.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT
XX hearts -
XX
XX Claim 1: SEQ ID NO 4442; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
CC
XX measuring human gene expression in a sample derived from human heart. The
CC
XX present sequence is one such probe. The probes may be used for
CC
XX predicting, measuring and displaying gene expression in samples derived
CC
XX from the human heart via microarrays. By measuring gene expression, the
CC
XX probes are useful for predicting, diagnosing, grading, staging,
CC
XX monitoring and prognosing diseases of the human heart and vascular system
CC
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC
XX congenital heart disease.
CC
XX Note: The sequence data for this patent did not form part of the printed
CC
XX specification, but was obtained in electronic format directly from WIPO
```

CC at ftp.wipo.int/pub/published_pcl_sequences.
XX Sequence: 1953 BP; 633 A; 316 C; 324 G; 680 T; 0 other;
SQ

Query Match 28.1%; Score 530.6; DB 22; Length 1953;
Best Local Similarity 99.3%; Pred. No. 6.5e-147;
Matches 533; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 280 aggtattcttgatgacgttggaagtgccgtatcatggaatcatctatgtatgga 339
    |||
Db 343 aggtattcttgatgacgttggaagtgccgtatcatggaatcatctatgtatgga 402
QY 340 gccctaaagccttagtgaacttgcttaactaaatgacataatgatacaagatgca 399
    |||
Db 403 gccctaaagccttagtgaacttgcttaactaaatgacataatgatacaagatgca 462
QY 400 ggaagtcactgtatggtgattggaagtgaagatttgcacaaatccttgacattcgac 459
    |||
Db 463 ggaagtcactgtatggtgattggaagtgaagatttgcacaaatccttgacattcgac 522
QY 460 ttattagatcggtatcatatgttgcataggaagtgaatccttaagtttgcttgaat 519
    |||
Db 523 ttattagatcggtatcatatgttgcataggaagtgaatccttaagtttgcttgaat 582
QY 520 ttcttcctatgttgatgacatgcatcatagaatgacatgctctacacaaataataa 579
    |||
Db 583 ttcttcctatgttgatgacatgcatcatagaatgacatgctctacacaaataataa 642
QY 580 tattgtgtgtatatacagaagaataatatacctccctgttggaactgaagatctgctg 639
    |||
Db 643 tattgtgtgtatatacagaagaataatatacctccctgttggaactgaagatctgctg 702
QY 640 tgggtaaacccctgattgacatgacataacatagatgaataaaccaataccagaatcca 699
    |||
Db 703 tgggtaaacccctgattgacatgacataacatagatgaataaaccaataccagaatcca 762
QY 700 atgctgaataatggcttcattatccagattccttgattgtcaagaatttaattgtg 759
    |||
Db 763 atgctgaataatggcttcattatccagattccttgattgtcaagaatttaattgtg 822
QY 760 ttcagcttgaggacactcaagttgagactaagatgacacggcaggtatataat 816
    |||
Db 823 ttcagcttgaggacactcaagttgagactaagatgacacggcaggtatataat 879
```

RESULT 13

AAK04514
ID. AAK04514 standard; DNA: 1953 BP.

XX AAK04514;
XX
DT 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe SEQ ID NO: 4505.

XX Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX epilepsy; cancer; ss.

OS Homo sapiens.

XX
PN WO200157275-A2.

XX
PD 09-AUG-2001.

XX
PF 30-JAN-2001; 2001MO-US00667.

XX
PR 04-FEB-2000; 2000US-0180312.

XX
PR 26-MAY-2000; 2000US-0207456.

XX
PR 30-JUN-2000; 2000US-0608408.

XX
PR 03-AUG-2000; 2000US-0632366.

XX
PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains -

XX Example 4; SEQ ID NO: 4505; 650bp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.

XX
SQ Sequence 1953 BP; 633 A; 316 C; 324 G; 680 T; 0 other;

Query Match 28.1%; Score 530.6; DB 22; Length 1953;
Best Local Similarity 99.3%; Pred. No. 6.5e-147;
Matches 533; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 280 aggtattcttgatgacgttggaagtgccgtatcatggaatcatctatgtatgga 339
    |||
Db 343 aggtattcttgatgacgttggaagtgccgtatcatggaatcatctatgtatgga 402
QY 340 gccctaaagccttagtgaacttgcttaactaaatgacataatgatacaagatgca 399
    |||
Db 403 gccctaaagccttagtgaacttgcttaactaaatgacataatgatacaagatgca 462
QY 400 ggaagtcactgtatggtgattggaagtgaagatttgcacaaatccttgacattcgac 459
    |||
Db 463 ggaagtcactgtatggtgattggaagtgaagatttgcacaaatccttgacattcgac 522
QY 460 ttattagatcggtatcatatgttgcataggaagtgaatccttaagtttgcttgaat 519
    |||
Db 523 ttattagatcggtatcatatgttgcataggaagtgaatccttaagtttgcttgaat 582
QY 520 ttcttcctatgttgatgacatgcatcatagaatgacatgctctacacaaataataa 579
    |||
Db 583 ttcttcctatgttgatgacatgcatcatagaatgacatgctctacacaaataataa 642
QY 580 tattgtgtgtatatacagaagaataatatacctccctgttggaactgaagatctgctg 639
    |||
Db 643 tattgtgtgtatatacagaagaataatatacctccctgttggaactgaagatctgctg 702
QY 640 tgggtaaacccctgattgacatgacataacatagatgaataaaccaataccagaatcca 699
    |||
Db 703 tgggtaaacccctgattgacatgacataacatagatgaataaaccaataccagaatcca 762
QY 700 atgctgaataatggcttcattatccagattccttgattgtcaagaatttaattgtg 759
    |||
Db 763 atgctgaataatggcttcattatccagattccttgattgtcaagaatttaattgtg 822
QY 760 ttcagcttgaggacactcaagttgagactaagatgacacggcaggtatataat 816
    |||
Db 823 ttcagcttgaggacactcaagttgagactaagatgacacggcaggtatataat 879
```

RESULT 14

AAK30016
ID AAK30016 standard; DNA: 1953 BP.

XX AAK30016;

XX
DT 06-NOV-2001 (first entry)

```
XX Human bone marrow expressed single exon probe SEQ ID NO: 4573.
DE
XX Human: bone marrow expressed exon: gene expression analysis; probe:
KM microarray; cancer; leukemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX MPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
XX Example 4; SEQ ID NO: 4573; 658bp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukemia and myeloma. The present sequence is one of
XX the probes of the invention.
XX
XX Sequence 1953 BP; 633 A; 316 C; 324 G; 680 T; 0 other;
XX
Query Match      28.1%; Score 530.6; DB 22; Length 1953;
Best Local Similarity 99.3%; Pred. No. 6.5e-147;
Matches 533; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 280 aggatattcttggtgattcttgggaagtcgcgtatcatggaatcaatctatgatggaa 339
DB 343 aggatattcttggtgattcttgggaagtcgcgtatcatggaatcaatctatgatggaa 402
QY 340 gccctaaggccttagtgaactgtttactaatgcatgaataatggtatcaagaatgcaa 399
DB 403 gccctaaggccttagtgaactgtttactaatgcatgaataatggtatcaagaatgcaa 462
QY 400 ggaagtcactgtagtggatgtggaagtggaatttcaccaatcttgaccatcgac 459
DB 463 ggaagtcactgtagtggatgtggaagtggaatttcaccaatcttgaccatcgac 522
QY 460 ttatagatgaggcatcatgtggtcataggaagtagaataatcctaaattgcttgaat 519
DB 523 ttatagatgaggcatcatgtggtcataggaagtagaataatcctaaattgcttgaat 582
QY 520 ttttccatggtggtgagatgtrcatcactcatgaagaatgctctcaaaaaaacaatata 579
DB 583 ttttccatggtggtgagatgtrcatcactcatgaagaatgctctcaaaaaaacaatata 642
QY 580 tatttggtctatacacaagaacatatacctccttggtggagccggaagacatctgtg 639
DB 643 tatttggtctatacacaagaacatatacctccttggtggagccggaagacatctgtg 702
QY 640 tgggtaaatcctgattgattgtgagcaataacatgaggaataaacgattccccaagaa 699
DB 640 tgggtaaatcctgattgattgtgagcaataacatgaggaataaacgattccccaagaa 699
```

```
DB 703 tgggtaaatcctgattgattgtgagcaataacatgaggaataaacgattccccaagaa 762
QY 700 atctgaataattggtctcatctccagattccttggtatgtrcaagaattatgtg 759
DB 763 atctgaataattggtctcatctccagattccttggtatgtrcaagaattatgtg 822
QY 760 tctcagcttggtgacctcaattaggaaccttaggaagatgccaagcggcaggttatata 816
DB 823 tctcagcttggtgacctcaattaggaaccttaggaagatgccaagcggcaggttatata 879

RESULT 15
AAIL4606
ID AAIL4606 standard; DNA; 1953 BP.
XX
AC AAIL4606;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #4539 for gene expression analysis in human cervical cell sample.
XX
XX Probe; human; microarray; gene expression; cervical epithelial cell;
XX cervical cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157278-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00670.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX MPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 25; SEQ ID NO 4539; 487bp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENPs). The present sequence is one such probe. The SENPs are derived
XX from human HeLa cells. The SENPs can be used to produce a single exon
XX microarray, which can be used for measuring human gene expression in a
XX sample derived from human cervical epithelial cells. By measuring gene
XX expression, the probes are therefore useful in grading and/or staging
XX of diseases of the cervix, notably cervical cancer.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1953 BP; 633 A; 316 C; 324 G; 680 T; 0 other;
XX
Query Match      28.1%; Score 530.6; DB 22; Length 1953;
Best Local Similarity 99.3%; Pred. No. 6.5e-147;
Matches 533; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 280 aggatattcttggtgattcttgggaagtcgcgtatcatggaatcaatctatgatggaa 339
DB 343 aggatattcttggtgattcttgggaagtcgcgtatcatggaatcaatctatgatggaa 402
```

```

QY 340 gccctaagagccctagtgaaactctgttacctaattgacataatggtatcaagaatgcga 399
    |||||||
Db 403 gccctaagagccctagtgaaactcttttactaatgacataatggtatcaagaatgcga 462
    |||||||
QY 400 ggaaggtcacctgtagtggttggaagtggagattttgccaaatcccttgaccattgcac 459
    |||||||
Db 463 ggaaggtcacctgtagtggttggaagtggagattttgccaaatcccttgaccattgcac 522
    |||||||
QY 460 ttattagatgcggtatcatgttggtcatagagaatgaagaatccctaagtcttctctgaat 519
    |||||||
Db 523 ttattagatgcggtatcatgttggtcatagagaatgaagaatccctaagtcttctctgaat 582
    |||||||
QY 520 ttttccctcatgtgtgtagatgtcacctcaatgaagaatgctctcaacaaataataaa 579
    |||||||
Db 583 ttttccctcatgtgtgtagatgtcacctcaatgaagaatgctctcaacaaataataaa 642
    |||||||
QY 580 tattgttgctatacagagaacataatacttcctctgttggaacctgagacatctgcttg 639
    |||||||
Db 643 tattgttgctatacagagaacataatacttcctctgttggaacctgagacatctgcttg 702
    |||||||
QY 640 tgggtaaaatccctgattgattgtgagcaataacatgaagataaacacagatcccaatcca 699
    |||||||
Db 703 tgggtaaaatccctgattgattgtgagcaataacatgaagataaacacagatcccaatcca 762
    |||||||
QY 700 atgctgaatatlttgctcattatltcccaattcttggattgttcaaggatttaagtgtg 759
    |||||||
Db 763 atgctgaatatlttgctcattatltcccaattcttggattgttcaaggatttaagtgtg 822
    |||||||
QY 760 tctcagcttgggcaacttcagttaggaactaagatgccaagccggcaggttatatat 816
    |||||||
Db 823 tctcagcttgggcaacttcagttaggaactaagatgccaagccggcaggttatatat 879
    |||||||

```

Search completed: May 9, 2002, 02:12:01
 Job time: 13401 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2002, 22:21:20 ; Search time 54.46 Seconds
(without alignments)
8529.006 Million cell updates/sec

Title: US-09-802-520-2
Perfect score: 1891
Sequence: 1 ggggaacgacgtgagtcgcg.....gtcaattatcgtggtga 1891

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTC05.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	521	27.6	521	US-09-323-873A-7	Sequence 7, Appli
2	476.4	25.2	1213	US-09-083-521-3	Sequence 3, Appli
3	425.8	22.5	444	US-09-030-607-214	Sequence 214, App
4	425.8	22.5	444	US-09-439-313-214	Sequence 214, App
5	363	19.2	366	US-09-030-607-215	Sequence 215, App
6	363	19.2	366	US-09-439-313-215	Sequence 215, App
7	310.2	16.4	322	US-09-323-873A-9	Sequence 9, Appli
8	306.8	16.2	328	US-09-030-607-212	Sequence 9, Appli
9	306.8	16.2	328	US-09-439-313-212	Sequence 212, App
10	290.4	15.4	1195	US-09-323-873A-1	Sequence 1, Appli
11	243.4	12.9	250	US-09-030-607-213	Sequence 213, App
12	243.4	12.9	250	US-09-439-313-213	Sequence 213, App
13	232.2	12.3	3627	US-09-323-873A-6	Sequence 6, Appli
14	125.4	6.6	592	US-09-439-313-342	Sequence 342, App
15	106	5.6	448	US-09-323-873A-11	Sequence 11, Appl
16	97.6	5.2	401	US-09-323-873A-12	Sequence 12, Appl
17	41.2	2.2	7218	US-08-232-463-14	Sequence 14, Appl
18	39	2.1	7218	US-08-232-463-14	Sequence 14, Appl
19	36.6	1.9	1219	PCR-US93-06251-11	Sequence 11, Appl
20	36.4	1.9	152331	US-09-128-155-16	Sequence 16, Appl
21	36	1.9	1474	PCR-US94-00545-19	Sequence 19, Appl
22	36	1.9	3478	US-08-396-479B-1	Sequence 1, Appli
23	36	1.9	3478	US-08-818-823-1	Sequence 1, Appli
24	35.6	1.9	684	US-08-998-416-672	Sequence 672, App
25	35.6	1.9	6453	US-08-306-691B-14	Sequence 14, Appl
26	35.6	1.9	6453	US-09-209-668-10	Sequence 10, Appl
27	35.6	1.9	6453	US-09-356-952-8	Sequence 8, Appli

28	34.8	1.8	4695	2	US-08-231-193A-57	Sequence 57, Appl
29	34.8	1.8	4695	3	US-08-486-273A-57	Sequence 57, Appl
30	34.8	1.8	4695	2	US-08-940-086A-57	Sequence 57, Appl
31	34.8	1.8	4695	4	US-08-940-035A-57	Sequence 57, Appl
32	34.6	1.8	5544	3	US-08-851-843A-68	Sequence 68, Appl
33	34.6	1.8	5544	4	US-08-974-549A-111	Sequence 111, App
34	34.6	1.8	5544	4	US-08-854-050-68	Sequence 68, Appl
35	34.6	1.8	5544	4	US-09-430-323-68	GENERAL INFORMA
36	34.6	1.8	5613	2	US-08-463-418-1	Sequence 1, Appli
37	34.2	1.8	1281	4	US-08-998-416-3	Sequence 17, Appl
38	34	1.8	669	1	US-08-463-115-17	Sequence 17, Appl
39	34	1.8	669	1	US-08-465-388-17	Sequence 17, Appl
40	34	1.8	2249	3	US-08-814-052-19	Sequence 19, Appl
41	34	1.8	2279	3	US-08-814-052-17	Sequence 17, Appl
42	34	1.8	2300	3	US-08-814-052-18	Sequence 18, Appl
43	34	1.8	3183	2	US-08-939-218A-1	Sequence 1, Appli
44	34	1.8	3187	5	PCR-US95-06815-1	Sequence 1, Appli
45	34	1.8	3192	1	US-08-706-037-26	Sequence 26, Appli

ALIGNMENTS

RESULT 1
US-09-323-873A-7
; Sequence 7, Application US/09323873A
; Patent No. 6329503
; GENERAL INFORMATION:
; APPLICANT: Daniel E. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Kahen Leong
; APPLICANT: Arthur B. Raitano
; APPLICANT: Douglas C. Saffran
; APPLICANT: Steve Chappell Mitchell
; TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
; FILE REFERENCE: 129,160S02
; CURRENT APPLICATION NUMBER: US/09/323,873A
; PRIOR FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/087,520
; PRIOR FILING DATE: 1998-06-01
; PRIOR APPLICATION NUMBER: 60/091,183
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 521
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-323-873A-7

Query Match 27.6%; Score 521; DB 4; Length 521;
Best Local Similarity 100.0%; Pred. No. 1e-148;
Matches 521; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1049 tgaatttcaaaatcccatagagatttgaaataaaccttccatagttccattt
Db 1 tgaatttcaaaatcccatagagatttgaaataaaccttccatagttccattt
QY 1109 ttgctctccatagatctacccgcaggtcttcgcgcagctgcttaccattatcagc 1168
Db 61 ttgctctccatagatctacccgcaggtcttcgcgcagctgcttaccattatcagc 120
QY 1169 caccaagatagagatttccaccttggttggaacctggttacaagtgtagaacaagct 1228
Db 121 caccaagatagagatttccaccttggttggaacctggttacaagtgtagaacaagct 180
QY 1229 tggattactaagttttttcttgctatggttccatggttgcttaagctcgttaccgat 1288
Db 181 tggattactaagttttttcttgctatggttccatggttgcttaagctcgttaccgat 240
QY 1289 gagaagtcagagagatattgttctcaacatgcttaccagcttaccatgcaaatat 1348

Accession	Sequence	Position
Dd	gaggaagcagagagatattgtttccacacatgctctatcagcaggttcctatcgaatat	300
Oy	1349 tgaaacctctggagatagagaagaagtcttgagaattgaatgatatctctttggcat	1400
Dd	301 tgaaacctctggagatagagaagaagtcttgagaattgaatgatatctctctggcat	360
Oy	1409 aatgagccttggcttacttcttcctcctcctgagacatctctatcctctcagtagaacaatgc	1466
Dd	361 aatgagccttggcttacttcttcctcctcctgagacatctctatcctctcagtagaacaatgc	420
Oy	1469 tttaaacctggagaaattcgaatttatacgtctacactctggatagatgtgcctgcgtcat	1528
Dd	421 tttaaacctggagaaattcgaatttatacgtctacactctggatagatgtgcctgcgtcat	480
Oy	1529 aagtaacttccatgttttaattatagatgagaagaagcgtt	1569
Dd	481 aagtaacttccatgttttaattatagatgagaagaagcgtt	521

```

1      RESULT      2
2      US-09-083-521-3
3      Sequence 3, Application US/09083521
4      Patent No. 6048970
5      GENERAL INFORMATION:
6      APPLICANT: Lal, Preeti
7      APPLICANT: Guegler, Karl J.
8      APPLICANT: Corley, Neil C.
9      TITLE OF INVENTION: PROSTATE GROWTH-ASSOCIATED MEMBRANE PROTEINS
10     NUMBER OF SEQUENCES: 7
11     CORRESPONDENCE ADDRESS:
12     ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
13     STREET: 3174 PORTER DRIVE
14     CITY: PALO ALTO
15     STATE: CALIFORNIA
16     COUNTRY: USA
17     ZIP: 94304
18     COMPUTER READABLE FORM:
19     MEDIUM TYPE: Floppy disk
20     COMPUTER: IBM PC compatible
21     OPERATING SYSTEM: PC-DOS/MS-DOS
22     SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
23     CURRENT APPLICATION DATA:
24     APPLICATION NUMBER: US/09/083,521
25     FILING DATE: Herewith
26     CLASSIFICATION:
27     ATTORNEY/AGENT INFORMATION:
28     NAME: CERRONE, MICHAEL C.
29     REGISTRATION NUMBER: 39,132
30     REFERENCE/DOCKET NUMBER: Pf-0527 US
31     TELECOMMUNICATION INFORMATION:
32     TELEPHONE: (650) 855-0555
33     TELEFAX: (650) 845-4166
34     * INFORMATION FOR SEQ ID NO: 3:
35     SEQUENCE CHARACTERISTICS:
36     LENGTH: 1213 base pairs
37     TYPE: nucleic acid
38     STRANDEDNESS: single
39     TOPOLOGY: linear
40     IMMEDIATE SOURCE:
41     LIBRARY: PROSTUT10
42     CLONE: 1691243
43     US-09-083-521-3

```

Query Match	25.2%;	Score 476.4;	DB 3;	Length 1213;
Best Local Similarity	99.8%;	Pred. No. 6.7e-135;		
Matches 477;	Conservative	0;	Mismatches 1;	Indels 0;
				Gaps 0;

Oy 1172 caagtatcggagatctccaccctggttggaaccctgttcaagtgtagaacaacagcttgg 1231
 |||||
 Db 1 CAAGTATAGGAGATTTCCACCTTGTTGGAAACCTGTTACAGTGTGAAACACAGCTTGG 60

OY	1232	atbctaagattttcttcgcataatgacatgctgcctcaacgcttcgttcacgaatg	1291
Db	61	ATTACTAAGTTTTCCTTCGCTATGCTCCATGTGGCTTACAGCTCTGCTTACCGATAG	120
OY	1292	aagtlcagaagagatatlttglttctcaacatgycgtltaacagcaggtlcalgcaaatltga	1351
Db	121	AAGGTCAAGAGATATTGTTTCTCAACATGCGTTATACAGAGTTCAATGAATATTGA	180
OY	1352	aaactcttggaatgaggaagaagtttgagaaatlgaatgatatctcttggcataat	1411
Db	181	AAACTCTTGGAATGAGGAAGAAGTTGGAAATGTAAATCTCCTTTGGCATAT	240
OY	1412	gagcttggcttacttccctcctctgagatcatctatcccttcagtagaagaatgctt	1471
Db	241	GAGCCTGGCTTACTTTCCTCCCTCGAGCACTTATTCCTCTCATGATAGAAATGATT	300
OY	1472	aaactggaagaatctcagttltaatctcagcttacaacttggaatgtagatgtgcgtctcataag	1531
Db	301	AAACTGGAGAAATTCAAGTTTATTCAGTTCACCTTGGATATGTGCGCTGCTATAG	360
OY	1532	tacttccatgltltaatltaatggaatgaaagagcttltggaagaagtactacagatt	1591
Db	361	TACTTCCATGTTTATTTATATGATGGAACAGCTTTTGAGGAAGATCACTACAGATT	420
OY	1592	ttataacccacaacttgccttctgtctgttggccctaattgtaacttgcgt	1649
Db	421	TTATACACACCAACTTGTCTGTGCTGTGTTTGGCCCTCAATGTATATCTGAT	478

```

: RESULT 3
: US-09-030-607-214
: Sequence 214, Application US/09030607
: Patent No. 6262245
: GENERAL INFORMATION:
: APPLICANT: Xu, Jiangchun
: APPLICANT: Dillon, David C.
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
: NUMBER OF SEQUENCES: 224
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SEED and BERRY LLP
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/030,607
: FILING DATE: 25-FEB-1998
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Makl, David J.
: REGISTRATION NUMBER: 31,392
: REFERENCE/DOCKET NUMBER: 210121.427C3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206) 682-6031
: INFORMATION FOR SEO ID NO: 214:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 444 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: US-09-030-607-214

```

Query Match	22.5%;	Score 425.8;	DB 4;	Length 444;
Best Local Similarity	98.6%;	Pred. NO. 8.7e-120;		

	Matches	438;	Conservative	0;	Mismatches	5;	Indels	1;	Gaps	1.
OY	688		accagaatccaatgctgaataatttgcttcaatatccagatctttgatgtcacaag						747	
Db	1		ACCAGAAATCCAAAGTCTAATAATTATTTGGGCTTCATTAATCCCAATTCCTTGATGTGCCAAG						60	
OY	748		gattaactgttcctcgaacttgggcacctcaagttagaacctaaggatgcacgcgcag						807	
Db	61		GATTTAAAGTGTCTCACACTTGGGCACCTTCAAGTTAGGAAGCTAAGATGCCACCGCAGG						120	
OY	808		ttaataatcgacgaacaataltcaacgcgcgaacaagttatlyaaacttgcgcag						867	
Db	121		TTTAATATATGACGACAAACAATATTCAAGCGCGCAACAAGTTATTAAGCTTGCCTCCCAAGT						180	
OY	868		tgaatttcatcccattgaacttggatctcttatcaatcagccgaagaagattgnaaatlac						927	
Db	181		TGAATTTATATCCCATGACTCTGGATGCTTAATATCAACCCMAAGAGATTGAAAATTTC						240	
OY	928		cacctcgaactttaactctctggagaagggccagltggtgtagctaataacttggccaca-						986	
Db	241		CCCTACGACACTTTTACTCTCTGAGAGAGGCGCAGTGGTAGCTTAAGAAGCTTGGGCACAT						300	
OY	987		tttttttcccttattcctcttggcagaatgyatattcatccatatgtcagaacaaccaag						1046	
Db	301		TTTTTTTTCTTAATTCCTTTGTCAGAAATGGATTCATCCATTAGCTANNAACCAACAG						360	
OY	1047		agtgcatttcaaaaattccctatagagatttgyaataaacttacctaatgttgcatt						1106	
Db	361		AGTAGCTTTTACAAAATTCCTATAGANNFTTGATATAAAACCTTACCTATACCTTGCATT						420	
OY	1107		aatttgtctctccctagtaactc						1130	
Db	421		ACTTTGGCTCTCCCTAATATTAACCTC						444	

```

RESULT      4
US-09-439-313-214
: Sequence 214, Application US/09439313
: Patent No. 6329505
: GENERAL INFORMATION:
:   APPLICANT: Xu, Jiangchun
:   APPLICANT: Dillon, Davin C.
:   APPLICANT: Mitcham, Jennifer L.
:   APPLICANT: Harlocker, Susan Louise
:   APPLICANT: Jiang Yugu
:   APPLICANT: Reed, Steven G.
:   APPLICANT: Kalos, Michael
:   APPLICANT: Fanger, Gary
:   APPLICANT: Retter, Mark
:   APPLICANT: Solk, John
:   APPLICANT: Day, Craig
:   TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
:   TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
:   FILE REFERENCE: 210121.427C9
:   CURRENT APPLICATION NUMBER: US/09/439, 313
:   CURRENT FILING DATE: 1999-11-12
:   NUMBER OF SEQ ID NOS: 575
:   SOFTWARE: FastSeq for Windows Version 3.0
:   SEQ ID NO 214
:   LENGTH: 444
:   TYPE: DNA
:   ORGANISM: Homo sapien
:   FEATURE:
:   NAME/KEY: misc_feature
:   LOCATION: (1)...(444)
:   OTHER INFORMATION: n = A,T,C or G
US-09-439-313-214

```

Query Match	22.5%;	Score 425.8;	DB 4;	Length 444;
Best Local Similarity	98.6%;	Pred. No. 8.7e-120;		
Matches 438; Conservative	0;	Mismatches 5;	Indels 1;	Gaps 1;

Qy	688	accagagatccaatgctcggaatattgtgtctcatattcccaagatcttgattgcgaag	747
Dp	1	accagagaaccaagtcggaatattgtgtctcatattcccaagatccttgattgcgaag	60
Qy	748	gattcaatgtgtctcagcgttggcacttaagttagagactaaggatgcgcgcgcag	807
Dp	61	gatttaatgtgtctcagcgttggcacttaagttagagactaaggatgcgcgcgcag	120
Qy	808	tttatatgacgaacaacatattccaagcgcgcgaacaaggattatgaacttgcgcgcagt	867
Dp	121	tttatatgacgaacaacatattccaagcgcgcgaacaaggattatgaacttgcgcgcagt	180
Qy	868	tgaattcatctcccatctggaacttggatctcatcatccagccagagagattgtaaatctac	927
Dp	181	tgaattcatctcccatctggaacttggatctcatcatccagccagagagattgtaaatctac	240
Qy	928	ccctaagactcttctatctctctgtgagagggccagttgtgttagctataaagcttgcacaa	986
Dp	241	ccctaagactcttctatctctctgtgagagggccagttgtgttagctataaagcttgcacaa	300
Qy	987	ttttttctcttatctctcttgcgaagatgattcatcccatatgctcgaacccaacag	1046
Dp	301	ttttttctcttatctctcttgcgaagatgattcatcccatatgctcgaacccaacag	360
Qy	1047	agtgactttacaacaaatctctatagagatgtgtaataaacccttaactaagtgtgcatt	1106
Dp	361	agtgactttacaacaaatctctatagagatgtgtaataaacccttaactaagtgtgcatt	420
Qy	1107	actttgctctccctagatataccgc	1130
Dp	421	actttgctctccctaaatataccctc	444

```

RESULT 5
US-09-030-607-215/C
: Sequence 215 Application US/09030607
: Patent No. 6262245
: GENERAL INFORMATION:
: APPLICANT: Xu, Jiangchun
: APPLICANT: Dillon, Davin C.
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
: NUMBER OF SEQUENCES: 224
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SEED AND BERRY LLP
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/030,607
: FILING DATE: 25-FEB-1998
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Makl, David J.
: REGISTRATION NUMBER: 31,392
: REFERENCE/DOCKET NUMBER: 210121.427C3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206) 682-6031
: INFORMATION FOR SEQ ID NO: 215:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 366 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: IS-09-030-607-215

```

Query Match	19.2%;	Score 363;	DB 4;	Length 366;
Best Local Similarity	99.2%;	Pred. No. 1e-100;		
Matches 363;	Conservative	0;	Mismatches 3;	Indels 0;
			Gaps	0;

Oy	1167	ggcccaagataagagaattccaccttggttgaaacctgtgttcacagtgtagaaaacag	1226
Db	366	ggcaccagatgtagaatttccacctctgggttgaaacctgggttACAGtgtagaaaacag	307
Oy	1227	cttgattactaagttttcttcttcgcatagtgtccatggttcacagccctcgtctaccg	1286
Db	306	CTTGATTACTAAAGTTTTTNTTGGCATGTCAGTGGCTACAGCCTCGCTTACCG	247
Oy	1287	atggaagaagtcagagagatatctgtctcacaatggtcttacagcaagttcatgcaaat	1346
Db	246	ATGGAAGAAGTCAGAGAAATTTGTTTCTTCACATGGCTTATCAGCAGGTCATCAAAAT	187
Oy	1347	attgaaaaccttgggaatctggagaagaatttggagaattggaattatatactcccttggc	1406
Db	186	ATTGAAAACCTCTGGAAATGAGGAAGAAAGTTTGAGAAATGAAATGATATNTCCCTTGGC	127
Oy	1407	ataatgagcccttggtcttaacttccctccgcagtcgaactctcatccctcagtggaat	1466
Db	126	ATAATGAGCCTTGCTTACTTTCCTTCCTGCGAGTACTCTTATCCCTTTCAGTAGCAAT	67
Oy	1467	gctttaaacttggagagaatcagtttatactcagctctacacttggatatagtcgtctgctc	1526
Db	66	GCTTTAAACGTGAGAGAAATTCAGTTTATTCAGTNTACACTTGGATATGTGCGCTCTGCTC	7
Oy	1527	ataagt 1532	
Db	6	ATAAGT 1	

RESULT 6
 US-09-439-313-215/C
 Sequence 215 Application US/09439313
 Patent No. 6329505
 GENERAL INFORMATION:
 APPLICANT: Xu, Jiangchun
 APPLICANT: Dillon, Davin C.
 APPLICANT: Mitcham, Jennifer L.
 APPLICANT: Harlockner, Susan Louise
 APPLICANT: Jiang Yuyui
 APPLICANT: Reed, Steven G.
 APPLICANT: Kalos, Michael
 APPLICANT: Fanger, Gary
 APPLICANT: Retter, Mark
 APPLICANT: Solk, John
 APPLICANT: Day, Craig
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
 TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
 FILE REFERENCE: 210121.427C9
 CURRENT APPLICATION NUMBER: US/09/439,313
 CURRENT FILING DATE: 1999-11-12
 NUMBER OF SEQ ID NOS: 575
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO: 215
 LENGTH: 366
 TYPE: DNA
 ORGANISM: Homo sapien
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: (1)...(366)
 OTHER INFORMATION: n = A,T,C or G
 US-09-439-313-215

Query Match	19.2%	Score 363;	DB 4;	Length 366;
Best Local Similarity	99.2%	Pred. No. 1e-100;		
Matches 363; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0

```

OY 1167 gaccacaagataagaagaattccacccttggttggaacctgttgacagtgtagaacaacag 1226
Db 366 GGCACCAAGGTAAGAGAAATTTCCACCCTGGTTGGAACACTGGTTACAGGTTAGAAAACAG 307
OY 1227 ctgagatactaagttttctctctgcgtatgcatagtgtgcacagcctctgtctacg 1286
Db 306 CTGGATTAAGTAAGTTTTTTTNTTCGGCTATGTCACATGTTGACCTTACAGCCTCTGCTTACCG 247
OY 1287 atgagaaggtcagaagaatatttgttcttcacaatgctctacagcaatgatacagaaat 1346
Db 246 ATGGAAGAGGTAGAGAAATATTTGTTCTCACATGCTTATCAGAGGTTTCATGCAAAAT 187
OY 1347 atbgaaaacctcttggaatlgagaagaagtttggagaattgaaatglatatcctcttggc 1406
Db 186 ATTGAAACCTCTTGGAATGAGGAAGAAGTTTGAGAAATGAATATATATNTCTTGGC 127
OY 1407 ataagagccttgcttaactcttcctccctcggaagtcactctatcccttcagtaggaat 1486
Db 126 ATAAATGACCTTGCTTACTTTCCTCTCGGACGTCATCTTATCTCTTCACTGATGACAT 67
OY 1467 gctttaaacttgagaagaatccagtttatccagtcctacacttgatctgcctctgc 1526
Db 66 GCTTTAAACGTGAGAGAAATTCAGTTTATTATTCAGTATACACTTGGAATATGTGCTCTGCTC 7
OY 1527 ataagt 1532
Db 6 ATAAGT 1
RESULT 7
US-09-323-873A-9
; Sequence 9, Application US/09323873A
; Patent No. 6329503
; GENERAL INFORMATION:
; APPLICANT: Daniel E. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Kahan Leong
; APPLICANT: Arthur B. Raffano
; APPLICANT: Douglas C. Saffran
; APPLICANT: Steve Chappell Mitchell
; TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
; FILE REFERENCE: 129.16USU2
; CURRENT APPLICATION NUMBER: US/09/323,873A
; CURRENT FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/087,520
; PRIOR FILING DATE: 1998-06-01
; PRIOR APPLICATION NUMBER: 60/091,183
; PRIOR FILING DATE: 1998-06-30
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 322
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-323-873A-9

```

```

RESULT 7
US-09-323-873A--9
; Sequence 9, Application US/09323873A
; Patent No. 6329503
; GENERAL INFORMATION:
; APPLICANT: Daniel E. Afar
; APPLICANT: Renee S. Hubert
; APPLICANT: Kahan Leong
; APPLICANT: Arthur B. Raitano
; APPLICANT: Douglas C. Salfran
; APPLICANT: Steve Chappell Mitchell
; TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
; TITLE OF INVENTION: EXPRESSED IN HUMAN CANCERS AND USES THEREOF
; FILE REFERENCE: 129.165US2
; CURRENT APPLICATION NUMBER: US/09/323, 873A
; CURRENT FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/087,520
; PRIOR FILING DATE: 1998-06-01
; PRIOR APPLICATION NUMBER: 60/091,183
; PRIOR FILING DATE: 1998-06-30
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 322
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-323-873A--9

```

Query Match	16.4%	Score 310.2	DB 4	Length 322
Best Local Similarity	99.0%	Pred. No. 1.1e-84		
Matches 312	Conservative 0	Mismatches 3	Indels 0	Gaps 0
Oy	991	ttttccttattcctcttgcagagatgtagatcatalcatalcgtagaaccacagatg	1050	
Db	8	ttttccttattcctcttgcagagatcgtatccatalcatalcgtagaaccacagatg	67	
Oy	1051	accttacaacattccatagagattggaataaaccttacctatgattgcattact	1110	
Db	68	accttacaacattccatagagattggaataaaccttacctatgattgcattact	127	
Oy	1111	tgctctccctgataacctcgcagagctctcttgcgcgtcgtattacaactttatggca	1170	
Db	128	tgctctccctgataacctcgcagagctctcttgcgcgtcgtattacaactttatggca	187	

RESULT 10

US-09-323-873A-1

Sequence 1, Application US/09323873A

Patent No. 6329503

GENERAL INFORMATION:

APPLICANT: Daniel E. Afar

APPLICANT: Rene S. Hubert

APPLICANT: Kahan Leong

APPLICANT: Arthur B. Raitano

APPLICANT: Douglas C. Saffran

APPLICANT: Steve Chappel Mitchell

TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS

FILE REFERENCE: 129.16USU2

CURRENT APPLICATION NUMBER: US/09/323,873A

CURRENT FILING DATE: 1999-06-01

PRIOR FILING DATE: 1998-06-01

PRIOR APPLICATION NUMBER: 60/091,183

PRIOR FILING DATE: 1998-06-30

NUMBER OF SEQ ID NOS: 32

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 1

LENGTH: 1195

TYPE: DNA

ORGANISM: Homo Sapiens

US-09-323-873A-1

Query Match 15.4%; Score 290.4; DB 4; Length 1195;
 Best Local Similarity 60.8%; Pred. No. 2.7e-78;
 Matches 474; Conservative 0; Mismatches 306; Indels 0; Gaps 0;

```

QY 935 acccttaccctctgagagagccagtggtgtagctatagcttggccattttt 994
DB 263 acccttccacagtggaacttcccaataaagctgtatataagctctgcatt 322
QY 995 ccttattcctctgacagagatgtatcatcatagtcagaaacagagtgact 1054
DB 323 tctttacacctctctgagaggaatgtaacaccttgcgaacttccatacaaat 382
QY 1055 ttacaataatcctatagagatgtgtaaaacactacattatgttgcatttgc 1114
DB 383 ttataaataatccaactgtgtcatcaacaaagcttgcgaatgttccatcctc 442
QY 1115 ctccctagttactcctgagagcttctctgagagctgtctacaaacttataag 1174
DB 443 ggcattggttaccctgacaggtgtgtaagacgaattgtccaaacttataag 502
QY 1175 gttatagagattccacctgtgttggaacattgttacaagttagaanaacag 1234
DB 503 gttataagaattccacattgttgaatgttgaatgttgaatgaagaagtttgc 562
QY 1235 actaagtttctctgcgtatgttgcattgttgcctacagcctctgcttacc 1294
DB 563 tctcagtttctcttctgttactgttactgtcattatagctgtcttaccac 622
QY 1295 gttcagagagatttttctcaacatgttctcagcaggttccatgaatattg 1354
DB 623 atcctcagatataaagttgttaactgttgatatacaacaggtccaaataaag 682
QY 1355 ccttgaagagagaaagtttggagaattgaaattgtatatccttggcctaag 1414
DB 683 tgcctgtatgagacagatgttggagaatgagattttagtctctctggaattg 742
QY 1415 ccttggcttacttccctctgagcagcttctatcccttgcagtgagcaattg 1474
DB 743 attggaataactgtctgttggctgtgacatattcatctcgttggatgacctt 802
QY 1475 ctggagagaatcagtttatttcaagttcactgtgatattgtcgtctgtcata 1534

```

```

DB 803 atgagagaatttccatataatcagagcaagcttagaattgttcccttactg 862
QY 1535 ttccacgtttaaattatgatgaaacagagctttagaagaagtagctacagatt 1594
DB 863 aatacagcattgtatttgcctcgtgaataagtgatgataataaataatttgat 922
QY 1595 tacacccaactgttcttgcctctgttgcctcctcaatgttaattctggt 1654
DB 923 tacacccaactttagtagctgttcttccctcaatgttgcctcgtatattaa 982
QY 1655 tattttcttccatgataagccgaagcctaaacgaatgaagaagagctggag 1714
DB 983 catactattcctgcacgtctgaggaagaactgaaagattagacatgttggag 1042

```

RESULT 11

US-09-030-607-23/c

Sequence 213, Application US/09030607

Patent No. 6262245

GENERAL INFORMATION:

APPLICANT: Xu, Jiangchun

APPLICANT: Dillon, David C.

TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS

NUMBER OF SEQUENCES: 224

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: WA

COUNTRY: USA

ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

CURRENT APPLICATION DATA: Patent Release #1.0, Version #1.30

APPLICATION NUMBER: US/09/030,607

FILING DATE: 25-FEB-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Makl, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.427C3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 213:

SEQUENCE CHARACTERISTICS:

LENGTH: 250 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-09-030-607-213

Query Match 12.9%; Score 243.4; DB 4; Length 250;
 Best Local Similarity 97.6%; Pred. No. 2e-64;
 Matches 244; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

QY 1283 accgatgagaagtcagagataattgttctcacaatgtctatcagcaggtcag 1342
DB 250 ACCGATGAGAAGTCAAGATATTTGTTNTTAACTGATGATCAGCAGTTATGC 191
QY 1343 aaatatgaaaactcttggaatgagaagaatttggagaattggaattgtatc 1402
DB 190 AAATATTGAAACCTCTTGATGAGGAAGATTGGAAGATTGAAATGATATNTCC 131
QY 1403 tggcataatagagcttgccttctcctcctcctggcagctactctatccctcag 1462
DB 130 TGGCATATAGAGCTTGCTTACTTTCTCTCTGCGAGTCACTTCTTCTCCCTTCA 71

```

QY 1463 caatgcttaactgagagaattcatttattcagctacattgataatgcctt 1522
|||||
Db 70 CAATGCTTTAACTGAGAGAAATTCAGTTTATTCAGTCTACACTGAGATGTCCTCT 11
QY 1523 gctcataagt 1532
|||||
Db 10 GCTCATTAAGT 1

RESULT 12

US-09-439-313-213/c
; Sequence 213, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yugu
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Rafter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 213
; LENGTH: 250
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(250)
; OTHER INFORMATION: n = A,T,C or G
US-09-439-313-213

Query Match 12.9%; Score 243.4; DB 4; Length 250;
Best Local Similarity 97.6%; Pred. No. 2e-64;
Matches 244; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1283 accgataggaagtcagagataattgtctcaacatgctatcaagagttcaatgc 1342
|||||
Db 250 ACCGATGAGAGAGTANAGAGATTTGTTTAAACATGATTCAGAGAGTTTCATGC 191
QY 1343 aatatgaaactcttggaatgaggaagaatttggaagaattgaaatgatatctctt 1402
|||||
Db 190 AAATTTGAAACTCTTGGAATGAGGAGAACTTGGAATTTGAAATTTATATTCCTT 131
QY 1403 tggcataatgagccttgctctctcctcctggaagtcacattatcccttcagtag 1462
|||||
Db 130 TGGCATTAATGAGCTTGGCTTACTTCCCTCGGAGACCTTCATCCCTCAAGTAG 71
QY 1463 caatgcttaactgagagaattcatttattcagctacattgataatgcctt 1522
|||||
Db 70 CAATGCTTTAACTGAGAGAAATTCAGTTTATTCAGTCTACACTGAGATGTCCTCT 11
QY 1523 gctcataagt 1532
|||||
Db 10 GCTCATTAAGT 1

RESULT 13

US-09-323-873A-6
; Sequence 6, Application US/09323873A
; Patent No. 6329503
; GENERAL INFORMATION:

; APPLICANT: Daniel E. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Kahn Leong
; APPLICANT: Arthur B. Raitano
; APPLICANT: Douglas C. Saffran
; APPLICANT: Steve Chappell Mitchell
; TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
; TITLE OF INVENTION: EXPRESSED IN HUMAN CANCERS AND USES THEREOF
; FILE REFERENCE: 129.16US02
; CURRENT APPLICATION NUMBER: US/09/323,873A
; CURRENT FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/087,520
; PRIOR FILING DATE: 1998-06-01
; PRIOR APPLICATION NUMBER: 60/091,183
; PRIOR FILING DATE: 1998-06-30
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 3627
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-323-873A-6

Query Match 12.3%; Score 232.2; DB 4; Length 3627;
Best Local Similarity 63.2%; Pred. No. 2.9e-60;
Matches 357; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

QY 935 actcttactctctgagagagccagtggtgtagctataaacttgccacatttttt 994
|||||
Db 293 actcttccacagtgagcacttgccacattaaatagctattatagcattctgcactt 352
QY 995 ccttattccttctgtagagatgtagtattcattcattgcttagaaccacagagtagct 1054
|||||
Db 353 tctttacactctctgtagagatgtagtattcattcattgcttagaaccacatttt 412
QY 1055 ttacaacattcctatagagattgtagaataaacttaccattagttgccaattcttgc 1114
|||||
Db 413 ttataaattccaatccttgtagcacaagaattctgccaatggttccatcactctct 472
QY 1115 ctcctagatlaactgcagagctctctgtagcagctgctatacaatttattagaccacaa 1174
|||||
Db 473 ggcattggttlaactgcagagctgtagagagcaattgccaacttccatlaattgaaaccaa 532
QY 1175 gtagaggaattccaccttggttggaaacccgtgtacaggttagaagaacagcttgatt 1234
|||||
Db 533 gataaagaattccacattggttggataagtagtataaagaagaagcttggtgct 592
QY 1235 actaaatttctctctgctatggtccatgttgccatacagcctctgcttaccgtagaag 1294
|||||
Db 593 tctcagttctcttctgctatggtccatgttgccatacagcctctgcttaccgtagaag 652
QY 1295 gtcagagagatattgttctcaacatgcttatacagcaggttcaatgcaaatatgaa 1354
|||||
Db 653 atccacagatacaagttgcttaaacatgagcagacatacaacaggtccacaacaataagaaga 712
QY 1355 cctctggaatgaggaagaatttggaagaattgaaatgatatcctcttggtcataatgag 1414
|||||
Db 713 tgcctggaatgagcatgattggttggagaattgagattatgctctctggaattgg 772
QY 1415 ccttgcttaactcctcctctgtagcagcactctatccctcagtaggaatgcttaaa 1474
|||||
Db 773 atggaataactgagctctgtagcagcactctatccctcagtaggaatgcttaaa 832
QY 1475 ctgagagaatcctgattattcag 1499
|||||
Db 833 atggaagaatcactatattcag 857

RESULT 14

US-09-439-313-342/c
; Sequence 342, Application US/09439313
; Patent No. 6329505

```

; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yugu
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 342
; LENGTH: 592
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-439-313-342

Query Match          6.6%; Score 125.4; DB 4; Length 592;
Best Local Similarity 61.7%; Pred. No. 2.8e-28;
Matches 198; Conservative 1; Mismatches 122; Indels 0; Gaps 0;

QY 935 acccttactctctgagagagccagtggtgtagctatagaacttgccacattttttt 994
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 321 ACCTCTTCACAGTCAGTGACCTTGCCAAATTAAGTGTCTATTATACATCTGTGACTTT 262
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 995 ccttattccttctgtagagagtgatgattatgcatatgctagaaccacaagagtgact 1054
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 261 TCTTTACACTCTCTGAGGAGATTAATTCACCCCTTAGCACTTCCTCATCAAAATTTT 202
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1055 ttacaactctctatagagatgtgaaataaacttaccatagttgccaatttctgtc 1114
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 201 TTATTAATATTCATCTCGTGTCAACAAAGTCTGCCAATGTTCCATCACTCTCT 142
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1115 ctccctagatatactcgcagtgctctgtagcagtgcttacaacttatttaagcacc 1174
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 141 GGCATTGGTTTACCTGCGAGTGTGATGAGCAGCAATGTGCAACTTCATATGGAACAA 82
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1175 gttagagagattccacactggttgaagaactggtacagtgtagaanaaacagcttgatt 1234
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 81 GATTAAGAACTTTCACATGTGGTGGTAAAGTGATGTTAACAAGAACARATGGGCT 22
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1235 actaagtttttcttcgcat 1255
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 21 TCTCAGTTCTTTTGTGCTGT 1
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
US-09-323-873A-11
; Sequence 11, Application US/09323873A
; Patent No. 6329503
; GENERAL INFORMATION:
; APPLICANT: Daniel E. Afar
; APPLICANT: Rene S. Hubert
; APPLICANT: Kahan Leong
; APPLICANT: Arthur B. Raitano
; APPLICANT: Douglas C. Saffran
; APPLICANT: Steve Chappell Mitchell
; TITLE OF INVENTION: NOVEL SERPENTINE TRANSMEMBRANE ANTIGENS
; FILE REFERENCE: 129,160SU2
; CURRENT APPLICATION NUMBER: US/09/323,873A
; CURRENT FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/087,520
; PRIOR FILING DATE: 1998-06-01
```

```

; PRIOR APPLICATION NUMBER: 60/091,183
; PRIOR FILING DATE: 1998-06-30
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 448
; TYPE: DNA
; ORGANISM: Homo Sapiens
; US-09-323-873A-11

Query Match          5.6%; Score 106; DB 4; Length 448;
Best Local Similarity 56.0%; Pred. No. 1.9e-22;
Matches 196; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

QY 1373 agttggagaattgaaatgatatctccttgccataatgacctggttacttccct 1432
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 30 agcctggtcagtagatcatatggtcttgggaataacttgggttttctgttctgact 89
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1433 cctggcagtaactctatccctcagtgagaactttaaactggagagaattcagtt 1492
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 90 ctgggaatacactctcttgccactgttagcaatgcaactggaagagatccgatt 149
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1493 taltcagttcaacttggatattgtctctgctccataagtaacttccatttattta 1552
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 150 tgtccagttccaacttgggtattttagccctgactctgtgtacagccacacccctgtgta 209
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1553 tggatgaaagagactttaggaagagtagctacagatttatacacaccacaacttgt 1612
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 210 cggltgggaagagattctcagcccttcaaatctcagatgtagtacttctcagcagc 269
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1613 tcttgctctgttcttgccctcaattgtaaltctgggtgaagattatttacttccatg 1672
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 270 gttagagcttatacttctctgcaactgtgctggtgatcaagttgtcctaactatgcatg 329
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1673 tataagccgaagactaaacgaattaaaagactggaagaagc 1718
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 330 tglagacaacaccccttacaagaagatccgcagagcttgggaagaagc 375
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: May 9, 2002, 02:06:15
Job time: 13495 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2002, 20:23:30 ; Search time 2036.92 seconds
(without alignments)
12530.066 Million cell updates/sec

Title: US-09-802-520-2

Perfect score: 1891

Sequence: 1 ggggaagcagcgtgagtcgcy.....gtcaattatcgtggttga 1891

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 segs, 6748477542 residues

Typal number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estrov:*
6: em_estrpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_estl2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	524.6	27.7	595	10	B1359597 384188 MA
2	451	23.8	561	10	B1133492 UI-M-BH3-
3	433.2	22.9	483	10	BF883142 QV3-ET021
4	411	21.7	712	12	A2428664 1M0212F14
5	377.8	20.0	2473	11	AK015015 Mus muscu
6	366	19.4	558	10	BM431438 B1105662
7	343.2	18.1	677	10	BM431438 1D016F03
8	326.6	17.3	948	10	BF784438 602892429
9	321.8	17.0	1287	11	BF784438 602110826
10	310.2	16.4	322	9	AA508880 ng86f03.s
11	275.4	14.6	573	10	AA508880 ng86f03.s
12	274.8	14.5	541	9	AA508880 ng86f03.s
13	272.8	14.4	1211	11	AA508880 ng86f03.s
14	262.2	13.9	680	9	AA508880 ng86f03.s
15	260.8	13.8	343	9	AA508880 ng86f03.s
16	250.6	13.3	600	10	AA508880 ng86f03.s
17	237.4	12.6	881	10	AA508880 ng86f03.s

18	235	12.4	506	9	A1930004
19	232.4	12.3	1035	9	AL555520
20	224	11.8	379	9	A1930312
21	211	11.2	625	10	BG712772
22	209.8	11.1	950	10	C23400
23	207.4	11.0	336	10	BF560333
24	199.6	10.6	691	10	BG919620
25	199	10.5	675	10	BE380978
26	194.4	10.3	444	9	BB859054
27	190.4	10.1	255	9	AA957846
28	186.4	9.9	710	10	BB875216
29	185.2	9.8	1003	10	B1551229
30	182	9.6	959	9	AL515863
31	180.6	9.6	242	9	AA997082
32	177.8	9.4	1073	9	AL577999
33	176.6	9.3	920	10	BB881257
34	176.2	9.2	654	10	B1066339
35	174.8	9.2	742	9	AU121735
36	174.2	9.2	734	10	BG923646
37	169	8.9	694	10	BG188296
38	169	8.9	796	10	BG188296
39	167	8.8	521	10	BF022394
40	166	8.8	600	9	A1686139
41	165.6	8.8	513	9	AA032221
42	164.4	8.7	961	10	BG762026
43	163.6	8.7	464	9	AW159609
44	158.6	8.4	756	10	B1753475
45	157.4	8.3	896	10	B1149488

ALIGNMENTS

RESULT 1
LOCUS B1359597 595 bp mRNA linear EST 01-AUG-2001
DEFINITION 384188 MARC 2P1G Sus scrofa CDNA 5', mRNA sequence.
ACCESSION B1359597
VERSION B1359597.1 GI:15055625
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Suidae.

REFERENCE 1 (bases 1 to 595)
Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
Stone, R.T., Heaton, M.P., Grose, W.M., Bennett, G.A., Laegreid, W.W.
and Keele, J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine

JOURNAL
COMMENT Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4360
Fax: 402 762 4390

FEATURES
source
1..595
/organism="Sus scrofa"
/db_xref="taxon:9833"
/clone_lib="MARC 2P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;

PCR Primers
FORWARD: AGGAACGCTATGACAT
BACKWARD: GTTTCACGACGACG
PLACE: 132 row: H column: 3
Seq primer: ATTAGTCACATATAG.
Location/Qualifiers

Library made from pooled tissue from testis, ovary, endometrium, hypothalamus, pituitary, and placenta.
 BASE COUNT 143 a 138 c 126 g 188 t
 ORIGIN

Query Match 27.7%; Score 524.6; DB 10; Length 595;
 Best Local Similarity 92.6%; Pred. No. 9,7e-123;
 Matches 551; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

OY 852 gaatttcgcccaggttgatattccattccattgagatccttatataccagca 911
 DB 1 GAATTCGCCGCGCATGGAATTCCTCCCTGACTGGGCTATATACATACAGCCAGG 60
 OY 912 gagattcaaaattaccctcagactcttactctctggaagagcgagtgatgact 971
 DB 61 GAGATTGAAATTTACCCCTCCGATTTACTCTCTGAGAGAGGCCAGTGTGTAGCC 120
 OY 972 ataaagttgscacatttttcccttattccttcttgcaagagatgcatcacaat 1031
 DB 121 ATAGCCGCGGCACATTTTCTTCTTATTCCTTTGCGAGATGATACATCATAT 180
 OY 1032 gctgaagaccacagagtgacttttacaataatcctatagagatgtgataaaacctta 1091
 DB 181 GCTGAAGAACCAACAGAGTGAATTTACAGATTCCTATGATGTGAATAAACCTTG 240
 OY 1092 cctatagttgcattacttctcctcctagatatactcagagtgcttcctgagagctgct 1151
 DB 241 CCCATCGTTGCGATTACTTTCCTCTGCTGCTACTGCGAGTCTCTCGCGAGCTGCT 300
 OY 1152 tatcaacttattagagcaccagatatagagattccaccttggttgaaacctgta 1211
 DB 301 TATACGCTTATATATGACACACCACTATAGCGATTCACCTTGCTGAGACCTGCTTA 360
 OY 1212 caggttgtaaaacagctgattacttaagttttcttcctgctcagctcgttgcctac 1271
 DB 361 CAGGTGTAAGAAACACTGGTGTACTCAGCTTTTCTTCGATCAGTCCATGTTGCTTAC 420
 OY 1272 agcctctgtaccagatgagaagtcagagatatgtttcttcaaatgagcttaccag 1331
 DB 421 AGCCTGCTGTACCCATGAGAGAGTACAGAGATACCTTTTTCATATGCGCTATACAG 480
 OY 1332 caggttataatgcaaatatgtaaaactcttggaatgagaagagtggttggaatgaaatg 1391
 DB 481 CAGGTTATATCAATATATGAAACTCTTGGAACAGAGAGAGTGTGAGATATGAATG 540
 OY 1392 tatatcccttctgcaatagagccttggttacttccctccctcgagcaatcact 1446
 DB 541 TATATCTCTTTGGCATATATGAGCTTGCTTACTTTCCTCTGCGAGTCACT 595

RESULT 2
 B1133492 561 bp mRNA linear EST 02-JUL-2001
 LOCUS B1133492/c
 DEFINITION UI-M-BH3-bix-g-01-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone
 VERSION B1133492
 ACCESSION B1133492.1 GI:14583740
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
 TITLE 1 (bases 1 to 561)
 JOURNAL Normalization and subtraction: two approaches to facilitate gene
 MEDLINE genome Res. 6 (9), 791-806 (1996)
 COMMENT
 Contact: Chin, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 20892-9643, USA

Tel: 301 443 1706
 Fax: 301 443 9890
 Email: MEST@mail.nih.gov
 The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the pineal glands tissue cDNA library Preparation: M.B. Soares Lab clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES

Source

Location/Qualifiers
 1..561
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-BH3-brx-g-01-0-UI"
 /clone_11b="NIH_BMAP_M_S4"
 /dev_stage="27-32 days"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pRT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The NIH_BMAP_M_S4 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged, normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH_BMAP_M_S4, NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1, NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library (NIH_BMAP_M_S4) was constructed as follows: PCR-amplified cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the NIH_BMAP_M_S4 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
 TAG_LIB=NIH_BMAP_M_S4
 TAG_TISSUE=pineal-glands
 TAG_SEQ=CACAC

BASE COUNT 154 a 115 c 129 g 163 t
 ORIGIN
 Query Match 23.8%; Score 451; DB 10; Length 561;
 Best Local Similarity 88.3%; Pred. No. 5.5e-104;
 Matches 490; Conservative 0; Mismatches 65; Indels 0; Gaps 0;
 OY 493 gtgaatccctaagttgctcgaattttcttcacatggtgagatgcatcaccatcatg 552
 DB 561 GCAGAAATCCCAAGTTTGCAATGAAATTTTCTCAGCGTAGAGCGTCAACCACAG 502
 OY 553 aagatgctcacaanaaataataatattgttctctacaacagagaaacattacact 612
 DB 501 AAGATGCTTTAAACAAAATAATATATTCCTGCTATACCAAGAACATTAACACT 442
 OY 613 cctctggagacccgagacatctgctgtgtaaaatcccgatggttgagaaataaa 672
 DB 441 CCTTGTGGAGCTTGAGACATCTGCTGTGGCAAAATCCCATTTGATGTGAGCAACACA 382

QY 673 tgaagataaaccagtaaccagaatccaatgctgaataatttgcttcaattatccagatt 732
|||||
Db 381 TGAGACTAATACAGATGCCAGATCCAGATGACAGAGTACCTGGCTCATATTATCCGACT 322
QY 733 ctctgattcacaaggaattaaatgctcagctgagcactcagtagaagcctaag 792
|||||
Db 321 CCTGATGTCATCAAGGATTTAATGATGATCTGAGCTTGAGGACACTTCAAGTAGGTCCTCAAG 262
QY 793 atgacgacgagcaggttatatatgacgaacaatatctaagggcagcaacaggttatg 852
|||||
Db 261 ATGCCAGCCGCCAGGTTTATATATGCAACAAATATCCAGCTCAGCAGCAGGTTATG 202
QY 853 aacttccgacagttgaattcattccattgacttgagctgagctcattatccagcagag 912
|||||
Db 201 AGCTGCCGCCAGCAGTGAATTTATTCCTGTTGACTTGAGGATCTTGTGTCAGCCAAAG 142
QY 913 agattgaanaattaccctcagactcttactctctgagagggccagtggtgtagcta 972
|||||
Db 141 AGATTGAAACCTTACCTGCGACTGTTACTCTGTGAGGAGGCGCAGTGTAGTACCA 82
QY 973 taagcttgacacatttttttcttcttcttcttcttcttcttcttcttcttcttct 1032
|||||
Db 81 TAAAGCTTGCCACATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 22
QY 1033 ctgaacaacacaga 1047
|||||
Db 21 CCAGAAAAA 7

RESULT 3
LOCUS BF883142 483 bp mRNA linear EST 17-JAN-2001
DEFINITION OV3-ET0211-071200-529-b10 ET0211 Homo sapiens CDNA, mRNA sequence.
ACCESSION BF883142
VERSION BF883142.1 GI:12273268.
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 483)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalha, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=QV3&t2=QV3-ET0211-071200-529-b10&t3=2000-12-07&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 26
High quality sequence stop: 483.
Location/Qualifiers
1. 483
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="ET0211"
/dev_stage="Adult"

/note="Organ: lung-tumor; Vector: puc18; Site: 1: Smal; Site: 2: SmaI; A mini-library was made by cloning products derived from ORESSES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 144 a 100 c 99 g 140 t
ORIGIN

Query Match 22.9%; Score 433.2; DB 10; Length 483;
Best Local Similarity 99.1%; Pred. No. 1.8e-99;
Matches 446; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 462 attagatggcgctatcatatgctgtaagaagaatccatgcttcttgaatt 521
|||||
Db 483 ATTAGATGGCGCTATCATATGTCATGAGAAATCTTAAGTTCTTCTGAAATT 424
QY 522 ttccctcatgtgtagatgtcaactcatcaagaatgctctcaaaaacaataata 581
|||||
Db 423 TTTCCTCATGTGTTAGATGTCTACTCATGAAAGATGCTCTCAAAAACAAATATATA 364
QY 582 ttgtgtctatacacagagaacattataccctcttggagacctgagacatctgtgtg 641
|||||
Db 363 TTGTGTCTATACACAGAGAACATTTACCTCCCTGTGGAGACTGAGACATCTGTTGTG 304
QY 642 ggttaaatccctgattgattgagcaataacagaggaataaaccagtaaccagaatccat 701
|||||
Db 303 GGTAAATCCCTGATGATGATGAGCAATTAACATGAGATTAACACAGTACCAGATCCAA 244
QY 702 gctgaatttgcttcatattcccaattcttattgttcaagaattatgtgtc 761
|||||
Db 243 GCTGAATTTTGGCTTCATTTATCCAGATTTCTTATATGTCAAGCATTTAATGTTGTC 184
QY 762 tcaagcttgacactcagttaggaacctaaagatgacagcagcaggaagttatataatgacg 821
|||||
Db 183 TCAGCTTGCGCACTTACGTAGGACCTTAAGATGCGCAGCGGAGGTTTATATATGACAG 124
QY 822 aacaatatcaagcgcagcaacaaggtatgtaactgcccagctgtaattatcc 881
|||||
Db 123 AACCAATTTCAAGCGCGCAACAGGTTATGAGCTTGCCGCGCAGTTGAATTTCAATCCC 64
QY 882 at-tgacttgagatccttcatcagccag 910
|||||
Db 63 ATGTGACTTGAGATCTTATATCAACAGAG 34

RESULT 4
LOCUS AZ428664 712 bp DNA linear GSS 03-OCT-2000
DEFINITION IM0212F14F Mouse 10kb plasmid U06C1M library Mus musculus genomic clone U06C1M0212F14 F, DNA sequence.
ACCESSION AZ428664
VERSION AZ428664.1 GI:10552677
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 712)
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausen, A. and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT 84112, USA

Db 121 TTTAAGAGAGTACTACAGGTTTATACACCAACCTTGTCTGCTGTTTG 180
QY 1629 cccctaatgttaattcttggaagaattatttattctccatgataaagccgaagcta 1688
Db 181 CCCCAATTGTAATTCGGGCAAGATGTTTACTCCCTCCATGATATAGCAAAAGTA 240
QY 1689 aaacgaattaaaaagctgggaaaaagaccattcttggaagaagtattgsgaaca 1748
Db 241 AAGGAATTAAAAAGGCTGGGAAAAGGCCAATTTCTAGAGAAGATATGGAGAGCA 300
QY 1749 attccatgtctcccccggagaaggtcaccaatgtgataataatggtttacagct 1808
Db 301 GTTCTCATCTCTACCGGAGAGGCTTACGTAATGATATAGATGCTCCCTGCT 360
QY 1809 gccatataagttctactatgcatatttatttgaactt---tacgtacgtacaag 1865
Db 361 GCTACATGAAGTCTCCGCAATGATATTTTATGACTTCTTACTTACTTGCAG 420
QY 1866 tatgtctcaaatatcgtggttga 1891
Db 421 TGCTCTGTCAAACTGCTGTGGCTGA 446

RESULT 7
LOCUS B1105662 677 bp mRNA linear EST 26-JUN-2001
DEFINITION 602892429F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5037414 5',
mRNA sequence.
ACCESSION B1105662
VERSION B1105662.1 GI:14556555
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 677)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLM1103 row: j column: 07
High quality sequence stop: 674.
Location/Qualifiers
1. 677

FEATURES
source
/organism="Mus musculus"
/strain="C2BECH II (feral)"
/db_xref="taxon:10090"
/clone="IMAGE:5037414"
/clone_lib="NCI_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary.
Stem cell origin."
/lab_host="DH10B"
/note="Organ: Lung; Vector: pCMV-SPORT6; Site: 1; Salt:
Site: 2; NotI: Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 177 a 172 c 160 g 168 t
ORIGIN

Query Match 18.1%; Score 343.2; DB 10; Length 677;
Best local Similarity 85.6%; Pred. No. 1.8e-76;
Matches 440; Conservative 0; Mismatches 68; Indels 6; Gaps 5;

QY 280 aggaattcttgatgacttggaagtgctcgtatcaatggaatcaatctctatgaggaa 339
Db 166 AGGTCAATTCCTGGGCTCTAGACGACTCTTACCAAGGAATCCATGATGAGAA 225
QY 340 gccctaaagagccttagtgaacttgctttacctaaatggaatgta-tcaaatgca 398
Db 226 GCCCTTAAGAGCCT--GGAGAGCTTTTGGCTTAATGGCATTAATGTAACCAAGAGCA 283
QY 399 aggaaggtcaactgtgagtgatgagtgagatttgccaaatcccttgaccattcga 458
Db 284 AGCGAGGTCAACGTTGGGGGTGATAGGAAGTGCGGATTTTGGCAAGTCTGTACCATTCGCG 343
QY 459 ctattagatgagctatcatatgctatgtagaagtgaatccttaagttgcttcgaa 518
Db 344 CTATTAAGATGTGCTACACGCTGTCATAGGAAGCAAGAAATCCCAAGTTTGATCGAA 403
QY 519 tttttctcactgtgtagatgctacatcatgaatgctctcacaataacata 578
Db 404 TTTTTCCTCACGTGTGAGACGTCAACCATTAAGATGCTTTAACAAAAAATAATA 463
QY 579 atattgtgctatcacagaagaacattatcctccctgtggaacctgagacatgctt 638
Db 464 ATATTGCGGTATCCATAGAGAACATTTACACCTCTTGCGGACCTTAGCATCTGCTT 523
QY 639 gtggtataaatcctgattgtagtgaacaataacatgaggaataaccagatcccgaa-tc 697
Db 524 GTGGGCAAAATCCATGATGATGAGAACACATGAGATGAACCAAGTCCAGAAATTC 583
QY 698 caatgcgaatattgcttcatcatcaccagattcttgatgtcaagaattaa-tg 756
Db 584 CAATGCAGAGTACTGCTGCTCATATTATCCCGAGACTCTTGATGTGTAAGGATTTAACTG 643
QY 757 ttgtcagcttg-ggcactcagttagacctga 789
Db 644 TGATCTCAGCTTGCGGCGCTTCACGTAGTCCCA 677

RESULT 8
LOCUS BF784438 948 bp mRNA linear EST 12-JAN-2001
DEFINITION 602110826F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4238935
5', mRNA sequence.
ACCESSION BF784438
VERSION BF784438.1 GI:12089474
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 948)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLM9851 row: d column: 08
High quality sequence stop: 660.
Location/Qualifiers
1. 948

FEATURES
source
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4238935"
/clone_lib="NCI_CGAP_Kid14"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: Kidney; Vector: pCMV-SPORT6; Site: 1; NotI:

BASE COUNT 238 a 395 c 358 g 296 t
 ORIGIN

Query Match 17.0%; Score 321.8; DB 11; Length 1287;
 Best Local Similarity 57.7%; Pred. No. 6.2e-71;
 Matches 634; Conservative 0; Mismatches 457; Indels 8; Gaps 3;

Qy 407 cactgtagtgatgtaggaatgtagatgtagcaatccctgagccattgactattag 466
 Db 120 CAAGTGGGCACTCTGGGCAATGGGGAATTTGCCGTTCCCTGGCCACACCGCTGGG 179
 Qy 467 atgcgcatcatatgtagcataggaagtagaactcctaagttgctctgaattttcc 526
 Db 180 CTCTGGCTTCAAGTGTGGTGGTGGGAGCCGCAACCCCAAGCAGCGCTGGCTCTCC 239
 Qy 527 tcatgtagtgatgtagcatcatatgaagatgctctcaaaaaaataataattgt 586
 Db 240 CTCTTAGCTCAAGTCACTTTCCAGGAGAACCGGTGAGCTTCCAGAGGTCACTTTGT 299
 Qy 587 tgcatacacagaacattatatactccctgtagacctgagacatctgcttggtgtaa 646
 Db 300 GCGCGTTCCTCGGAGACATATCTCTACCTGTGCACTGCTGCTGACCAAGTTGGCTGCCAA 359
 Qy 647 aatcctgattgtagtgcaataacatgagataa-----ccagtaaccagaatccaa 700
 Db 360 GATCTCGTGGATGTAGTAACCAACCCACGAGAAAGAGACATCTTCAGACACCGCCAGCTGTA 419
 Qy 701 tgcgaatatttgtagctcatatattccagatcttctgtagtcaagaagattatgtgt 760
 Db 420 CGCGATACCTGGCTGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 479
 Qy 761 ctcaagcttgagcaactgagtagaagacatgagtagccagccagaggttatatatcag 820
 Db 480 CTCTGATGGGCGCTTACAGGCTGGCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 539
 Qy 821 caacaataatcaagcgagcaacaaggttatgtaactgtagccgagtagtgaattcattcc 880
 Db 540 TGATCACCCAGAACGCCAAGGAGCACATCTGAGATGGACGCCCATGGGTTTACACACC 599
 Qy 881 catgtagtgtagtctcatcatcatcaacagagagatgtaaaattaccctagacacct 940
 Db 600 CTGGAGATGGATCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 659
 Qy 941 tactctctgagagagcgcaagtgtgtgtagtataaagcttgtagccacatttttccctta 1000
 Db 660 TCCATCCTGGAAGGTGCCACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 719
 Qy 1001 ttccttctgtagagatgtagtattcatcatatgctagaacaacagagtagactttcaaa 1060
 Db 720 CAACCTTCATCCGAGACCTTCTACAGCCTATCCGAAAGATGAGAACAAAGTTCTCAAA 779
 Qy 1061 aattccatagagatgtagtataaaaccttaccatagtgtagtacttctgctccct 1120
 Db 780 GATGCCCTTCTGTGTGCTACACACACATACCTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 839
 Qy 1121 agtataacctgtagctctctgtagcagctgtagtatacaacttattagcgagcaagtag 1180
 Db 840 AGGTGACCTGCGGCTGT 898
 Qy 1181 gagattcgaaccttgtagtgaacacctgtagttagaacaacagcttgtagtacttaag 1240
 Db 899 GCGCTTCCCGAGCTGT 958
 Qy 1241 ttttctctgtagttagttagttagttagttagttagttagttagttagttagttagttag 1300
 Db 959 CT 1017
 Qy 1301 gagaatattgttctcaacaatgtagttagttagttagttagttagttagttagttagttagt 1360
 Db 1018 CCCTTACGACCTGT 1077

Qy 1361 gaatgaggaagattgtagaattgaatgtagtattcttcttgtagtcaatgagccttg 1420
 Db 1078 GGCTGAGGAGAACTGCTGAGGATGAGATATACCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1137
 Qy 1421 cttaattccctctgtagcagtagttagttagttagttagttagttagttagttagttagttag 1480
 Db 1138 CATTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1197
 Qy 1481 agaattcagttattcag 1499
 Db 1198 GGAATTCAGCTGTGTGTGAG 1216

RESULT 10
 AA508880 322 bp mRNA linear EST 13-AUG-1997
 LOCUS
 DEFINITION
 n986f03.s1 NCI_CGAP_Pr6 Homo sapiens CDNA clone IMAGE:941693, mRNA
 sequence.
 ACCESSION
 AA508880 GI:2245821
 VERSION
 AA508880.1
 KEYWORDS
 EST.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
 1 (bases 1 to 322)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: c9abps-r@mail.nih.gov
 Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuagui,
 M.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: David B. Krizman, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/BLN at:
www-bio.lnbl.gov/bdrrp/image/image.html
 Insert Length: 1000 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 249.
 Location/Qualifiers

FEATURES
 source

1..322
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:941693"
 /clone_lib="NCI_CGAP_Pr6"
 /sex="male"
 /tissue_type="prostate"
 /lab_host="DH10B"
 /note="Vector: PAMP10; mRNA made from prostatic
 intraepithelial neoplasia (low-grade), CDNA made by
 oligo-dT priming. Non-directionally cloned.
 Size selected on agarose gel, average insert size 600 bp.
 Reference: Krizman et al. (1996) Cancer Research
 56:5380-5383."

BASE COUNT 84 a 70 c 62 g 106 t
 ORIGIN

Query Match 16.4%; Score 310.2; DB 9; Length 322;
 Best Local Similarity 99.0%; Pred. No. 3.6e-68;
 Matches 312; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 991 ttttcttattctcttctgtagagatgtagtattcatcatatgctagaagaacagagag 1050
 Db 8 TTTTCCCTTAATCTTGTGACAGATCTGATTCATATGCTAGAAACCAACAGAGTG 67
 Qy 1051 accttcaaatctccatagatgtagtgaataaaccttactatagttgtagtactt 1110
 Db 68 ACTTTTACAAATTCCTATAGAGATTGTGATTAATAACCTTACTATAGTTGCTCACTTATT 127

Qy 1111 tgcctccctagctacacgcaggtcttcgtgcagctgcttatcaacttattacgca 1170
|||||
Db 128 tgcctccctagctacacgcaggtcttcgtgcagctgcttatcaacttattacgca 187
|||||
Qy 1171 ccaagatagagagattccaccttggttgaaacctggttaccagtgtgaanaaacagcttg 1230
|||||
Db 188 CCAAGTATAGAGATTCCACCTTGTTGAAACCTGGTTACAGTGTAGAAAACAGCTTG 247
|||||
Qy 1231 gattactaagttttcttcgtctatgtgtccatgtgtgctacagcctctgcttaccgata 1290
|||||
Db 248 GATTCTAGTGTCTTCCTGCTATGTCATGTCCTACAGCTCTGCTTACCGATGA 307
|||||
Qy 1291 gaagtcagagagat 1305
|||||
Db 308 GAGGTCAGAGAGAT 322
|||||
RESULT 11
BM491178 573 bp mRNA linear EST 07-FEB-2002
LACUS
DEFINITION ppg2n.pk005.18 Normalized Chicken Pituitary/Hypothalamus/Pineal
Library (ppg2n) Gallus gallus cDNA clone ppg2n.pk005.18 5' similar
to gb|AA50539.1 (A1029586) dudulin 2 [Mus musculus], mRNA
sequence.
ACCESSION BM491178
VERSION BM491178
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 573)
Porter,T.E. and Cogburn,L.A.
ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA
library, USDA/IRAFs Animal Genome Project
Unpublished (2002)
JOURNAL Contact: Larry A. Cogburn
COMMENT University of Delaware
Townsend Hall, Newark, DE 19717, USA
Tel: 302-831-1335
Fax: 302-831-2822
Email: cogburnudel.edu, www.chickest.udel.edu.
FEATURES
Source
1. 573
/organism="Gallus gallus"
/strain="Commercial broiler chickens"
/db_xref="taxon:9031"
/clone="ppg2n.pk005.18"
Pituitary/Hypothalamus/Pineal Library (ppg2n)"
/sex="Male and female"
/tissue_type="Pituitary Gland/Hypothalamus/Pineal Gland"
/dev_stage="Embryonic (d12,d14,d19); post-hatch (1,3,5,7,9
weeks)"
/lab_host="E. coli EMDH10B"
/note="Vector: pCMVSPORT6; Library made from equivalent
pools of total RNA isolated from each tissue at different
ages. Single pass sequencing from 5'-end"
BASE COUNT 150 a 153 c 144 g 126 t
ORIGIN

Db 212 AAACCTTCTACCAATATGCGCCAAATGGCATCAGAGATCCAGTAGATCAGTAGGCA 271
|||||
Qy 418 tgattggaagtgagatttcccaaatcccttgacatccttgacttattagatgcgtatc 477
|||||
Db 272 TCATTGGAGGTGAGAGCTTTGCAAAATCATTCATGACCTTTGGTGTAGATGTGGTAC 331
|||||
Qy 478 atgtgtgctaggaagtagaataatccaaagtgtgcttcgtcaattttcttcctcagtgtg 537
|||||
Db 332 ATGTTGTTTAGAGACCAAAACCCAAACTGTGTCAGAGTTCCTTCCCATGTGTG 391
|||||
Qy 538 atgtcatcatatgaagaatgctcttcacaaacaaataataattgtgtctacaca 597
|||||
Db 392 ACGTCACATACCATGATGAGATGGGTACCAAAACAAACATTTTGTAGCATATTCATA 451
|||||
Qy 598 gagaacattatacctccctgtgtgagcttgagacatcctgcttgggttaaacctgattg 657
|||||
Db 452 GAGAACACTACACGCTTTGTGGACCTTCACACATTTACTCGGTAAAAATTCGTGTG 511
|||||
Qy 658 atgtgagcaataacatgagataaaccaggtaccacgaatccaatgtgataattgctt 717
|||||
Db 512 ATGTCAGACACATATACAGAGTAGACCATATCCAGACTCCATGAGATTTTGGCAT 571
|||||
Qy 718 c 718
Db 572 c 572
RESULT 12
A1747886 541 bp mRNA linear EST 22-JUN-1999
LOCUS
DEFINITION u103ell.y1 Sugano mouse kidney mklia Mus musculus cDNA clone
IMAGE:2065004 5', mRNA sequence.
ACCESSION A1747886
VERSION A1747886
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 541)
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk,R., Ritter
E., Kohn,S., Shin,T., Jackson,T., Gardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
JOURNAL Contact: Marra M/WashU-NCI Mouse EST Project 1999
COMMENT Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGT:994192
Seq primer: custom primer used
High quality sequence stop: 498.
FEATURES
Source
1. 541
/organism="Mus musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:2065004"
/clone_lib="Sugano mouse kidney mklia"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: Kidney; Vector: pME18S-FL3; Site_1: DraIII
(CACTGCTG); Site_2: DraIII (CACTCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
(ATGTGGCCCTTTTATTTTATTTT); double-stranded cDNA was


```

evidence:ISS
putative
six transmembrane epithelial antigen of the prostate"
/codon_start=1
/protein_id="BAB26938.1"
/db_xref="gi:12845881"
/translation="MEISDVNPEQLMKMKPKGNLEDDSYKSDGETSMKLRGLS
HLOAVHVDADCESELOHTEFFENMRPLPKVAIISLFEVLREIPLYVTSR
EYFVKIPIIVINKEPVAITLAIYLPGLIAVVOIRNGSTYKRPPLDPMELA
KKQGLISFFPVAIVHAYSSYSPHRSRIRKLKWAYKOVONKEDANVEDVHMEI
YVSLGIVGLATLALAVTSPISVSDSLWRFHYIOSKLGIVSLGLGVHAFVAMK
WVDYSQFVWYMPFPMIAVFLPLVLICKIALCPLEKRLIKIRCGMEDVSKINRTE
MASRL"
polyA_signal
1185..1191
/note="putative"
polyA_site
1211
/note="putative"
BASE COUNT      306 a      275 c      293 g      337 t
ORIGIN

```

```

Query Match      14.4%: Score 272.8; DB 11; Length 1211;
Best Local Similarity 59.4%: Pred. No. 1.9e-58;
Matches 463; Conservative 0; Mismatches 317; Indels 0; Gaps 0;

```

```

QY 935 actcttactctcttgagagagcgagtggtgtagctataagcttggccacattttttt 994
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 295 ATTCCTTCCAAACGTGGCGCTTGCAGTAGAAGTCCCTGCATCATATCTCCCTGACTTT 354
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 995 ccttattcctcttgacagagatgtagtcatcatagcttagaaccacagagtgactt 1054
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 335 CCTGTACACACTTGTGAGGGAATCATCCCGTTAGTAACCTTCCCGTAACATATTT 414
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1055 ttacaaatccctatagagatgtgtaataaaaccttacctatagttgcattacttgc 1114
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 415 TTATATAATTCCAAATCTGTGTATTACAAAGCTTTGCCAATGTGCGCATTAACCTCTT 474
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1115 ctccctatataccctgcagagctctctgcagctgctatacaattataagaccaca 1174
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 475 GGCACCTGGCTATTTTCCAGAGAGAGCTTAGCGGCACTTGTACAGCTTCGCAATGGAACCA 534
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1175 gtatagagattccacacttggttgaaacctggttacagtgtagaanaacagcttgatt 1234
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 535 GTACAGAAGATTCCACCCCTGTTAGATAGATGATGCGGCCAGAGAAGCACTTTGGTCT 594
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1235 actaagttttcttcctgcatagtgctacagctgcacagcctgcctaacagatgagaag 1294
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 595 CCTCAGCTTCTTTTGGCTGCTGCTACAGCTGTTTACAGTCTCTCATACCCCAATGAGAG 654
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1295 gtcaagagagatatgtttctcaacatggtctatcacagagcttagtcaaatatgaaaa 1354
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 655 ATCTTACAGATACAGACTCACTCACTGAGCTTACAAAGCGTTCAACAAACAAAGAGCA 714
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1355 ctcttgagatgagagaagatttgtagaataatgtatatactcctcttgcatataag 1414
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 715 TGCCGTGGGTAAGAGCATGTGCTGAGAAATGAGATTATGTCCTCCGCGGATGTGGG 774
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1415 ccttggttactcttccctcctgcagctcaactctatcccttcagtgcagcaatgtcttaa 1474
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 775 GCTGGCCACTCTTGGCTCTCTGCTGCTGACATCTATTCACCTGTGAGGAGCACTTTTAA 834
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1475 ctgagagaagaattcagttatcatcagctacacatgatatgctcgtctgcatataagtc 1534
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 835 CTGGAGAGAAATTTCACTATATTTCAGAGCAAACTAGGAATGTCTCTCTCTGCGGAC 894
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1535 ttccatgttttaattttagatgagaacagagcttctgagagaagtagctacacattta 1594
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 895 AGTACACGCTTTGGTTTGGCTGGAATAAATGGTAGATGTGCATCAATTTGTAATGTA 954
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1595 taacaccacaacttgctctgctctgcttgcctcccaatgtaattctgggtaagat 1654
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 955 CATGCCCTCGACTTTATGATAGTGTGTTCCCTCCAACTCTTGTCTCATATGTAAAT 1014
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

QY 1655 tatttattccttcacatgataagccgaagctaaacgaatlaaaaaagctgggaaaa 1714
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1015 CCCCCGTGCTCCCTGCTGCTGCTAGAGAGAAGATACGAGATAGATGTGCTTGGGAAGA 1074
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 14
BB603988
LOCUS
DEFINITION BB603988 Riken full-length enriched, 15 days embryo head Mus
musculus cDNA clone D930007L06 5', mRNA sequence.
ACCESSION BB603988
VERSION BB603988.2 GI:16450725
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.

```

```

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 680)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Osaki
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.

```

```

RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
On Dec 5, 2000 this sequence version replaced gi:11555390.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.c.riken.go.jp,
URL: http://genome-gsc.riken.go.jp/

```

```

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagii,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multichipillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanka,I., Aizawa
,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome-gsc.riken.go.jp/) for
further details.

```

```

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Location/Qualifiers
1..680
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="D930007L06"
/clone_lib="RIKEN full-length enriched, 15 days embryo

```

FEATURES

```

source

```


Thu May 9 08:44:32 2002

us-09-802-520-2.rst

Page 13

Db	103	CCCTGTGGACCTGAGGCA	TCTGCTGGTGGGCAAAAGT	CTCTATTGATGAACAAACAA	44
Qy	673	tgaggataaacagcagcagcagatccaa	700		
Db	43	TGAGAGTAAACCGATACCCAGATCCAA	16		

Search completed: May 9, 2002, 00:02:57
Job time: 13167 sec

THIS PAGE BLANK (USPTO)